DOCUMENT RESUME

ED 266 159 TM 860 102

AUTHOR	Herman, Joan; Winters, Lynn
TITLE	Test Design Manual: Guidelines for Developing Diagnostic Tests. Diagnostic Testing Project.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Dec 85
GRANT	NIE-G-83-0001
NOTE	99p.; Portions of appendices contain small print. Some graphics may not reproduce clearly.
PUB TYPE	Guides - Classroom Use - Guides (For Teachers) (052) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Diagnostic Tests; Elementary Secondary Education; Field Tests; Instructional Development; Skill Analysis; Task Analysis; *Teacher Made Tests; *Test Construction; Test Format; *Test Items
IDENTIFIERS	Diagnostic Testing Project

ABSTRACT

This manual is designed to provide teachers with technically sound, easy-to-follow procedures for developing diagnostic tests to be used with their own students. The definition of diagnostic testing is broadened to include any tests systematically designed to provide information about skills that students have or have not mastered. A five-step test development process divides the manual into five sections: (1) Specify the Diagnostic Skills to be Tested; (2) Create the Test Description; (3) Write the Test Items; (4) Review Test Items; and (5) Try Out Items. Much space is devoted to the conceptual plan underlying the diagnostic test, the skill map. Several strategies for developing the skill map through task analysis are mentioned. The application of research-based analytic methods to the description of important skills is treated as the core of the description section. Once the conceptual base for the diagnostic test is documented in the skill map and the test description, the steps for writing, reviewing, and trying out diagnostic items remain much the same as those used in survey testing programs. Appendices include model test descriptions, many with graphics, a test description review form, sample constructed response scoring guidelines and sample test item formats. (LMO)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED266159

DELIVERABLE - DECEMBER 1985

DIAGNOSTIC TESTING PROJECT

Test Design Manual:

Guidelines for Developing Diagnostic Tests

Study Director
Joan Herman

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. Griffith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Grant Number
NIE-G-83-0001

Center for the Study of Evaluation
Graduate School of Education
University of California, Los Angeles

TM 860 102

This report was authored by
Joan Herman & Lynn Winters

This project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

INTRODUCTION AND OVERVIEW

Help At Last

Teachers know that diagnostic tests have an important place in planning instruction and improving student learning. Unlike the more global standardized achievement tests administered yearly in most school districts, diagnostic tests provide profiles of strengths and weaknesses for specific students on a limited number of clearly defined instructional objectives.

The narrow focus and curricular relevance of diagnostic tests make them a useful tool for instructional planning throughout the school year:

o In the fall, teachers use diagnostic tests to determine student or class mastery of prerequisite skills. Individual and class needs are pinpointed so that the teacher can target instruction at the appropriate level. Sometimes test results indicate that students can be effectively grouped for instruction on particular objectives; other times, diagnostic tests reveal that an entire class may need to review certain material or can be skipped ahead of where the teacher traditionally begins instruction.

o During the year, diagnostic tests help teachers make regrouping decisions and identify students who may need extra help or more challenging assignments. They keep the teacher informed about class progress toward mastering grade level content and skills.

o At the end of particular instructional units, a semester or even a year, teachers may use diagnostic tests to judge the effectiveness of their instructional materials or methods.

Although teachers know that diagnostic tests can enhance instructional planning and evaluation, such tests are difficult to find. Very few commercially published tests are diagnostic. Those tests which are labelled diagnostic, such as the Prescriptive Reading Inventory (CTB-McGraw Hill) or the Stanford Diagnostic Reading Test (Psychological Corporation), don't always relate to a particular teacher's grade level scope and sequence. Published diagnostic tests tied to a particular district's instructional program in science, social studies, art, music, literature, writing, and oral language simply don't exist.

Of course, many teachers have attempted to solve the problem in the same manner as teachers have traditionally coped with a lack of materials for their classrooms: they have developed their own tests and methods for obtaining diagnostic information. And they have created these tests with little support or technical guidance from testing experts.

Guidelines for Developing Diagnostic Tests is designed to rectify that omission and to provide teachers with technically sound and easy to follow procedures for developing diagnostic tests for use with their own students. Guidelines will help teachers to design better tests, write better test items, and make better decisions about which students have really mastered materials and which need further instruction.

Diagnostic Tests

Most teachers think of diagnostic tests as tests specifically designed to measure a set of narrowly-defined skills in order to a provide profile of strengths and weaknesses for an individual pupil. This manual broadens the definition of diagnostic testing to include any tests systematically

designed to provide information about skills that students have or have not mastered. Diagnostic testing, then, includes tests used to make instructional decisions about groups of students as well as about individuals.

The test development strategies described in this manual are appropriate for creating any type of test. However, the manual focuses on assessment of academic achievement. Although non-cognitive factors such as student attitude, motivation, nutrition, and vision must be included in a comprehensive diagnostic system, they are beyond the scope of this manual. Because a thorough understanding of which academic skills students have or have not acquired is at the heart of a sound diagnostic approach to instructional planning, the procedures and examples presented relate specifically to the development of achievement tests.

Five Steps to Better Diagnostic Tests

How can teachers develop their own diagnostic measures? How do they improve tests they have already created? A five step test development process can be used to develop new tests or to improve existing teacher-developed tests. The process is straightforward and employs procedures that teachers already use in some form or another to plan instruction and create their own instructional materials. The figure 1 compares this five step process to steps teachers follow in preparing materials for their classes.

Each step in the diagnostic test development process is briefly described below and summarized in figure 2. The Guidelines are divided into five sections corresponding to the five steps in the process. Each

FIGURE 1

How Test Development and Instructional Planning Are Alike

Steps in Instructional Planning	Steps in Test Development
1. Identify student needs o identify where students are with relation to curricular goals o select objectives and activities that will enable students to reach goals o make rough teaching timeline	1. Specify skill map o identify important skills for which diagnosis is needed o do task analysis to identify skills that build toward test o sequence subskills from simple to complex
2. Create Lesson Plan o select skills from timeline prior to lesson o review grade level scope and sequence o consider adopted texts and other materials available o write lesson objective and method for checking whether students have mastered it	2. Create Test Description o identify skills to test o specify content to be covered on test o select item format
3. Develop/Gather Materials	3. Write Test Items
4. Select Best Materials	4. Review Test Items
5. Give Lesson and Adjust to Student Needs	5. Try Out Items and Revise

FIGURE 2

The 5 Step Diagnostic Test Development Process

Specify Skill Map
- o Identify skills leading up to those being tested
- o Sequence skills in the order to be taught

Create Test Description
- o Select skills and content from skill map
- o Clarify skills
- o Select item format

Write Test Items
- o Draft items following specifications contained in test description
- o Attend to item writing conventions for item type selected

Review Test Items
- o Check against test description
- o Check for technical flaws
- o Check wording for clarity, unnecessary ambiguity/complexity

Try Out Items
- o Use tryout information to
  1. identify poor items
  2. adjust test description
  3. doublecheck completeness and sequence of list of prerequisites
- o Determine number of items on test

section provides a full description of the step, presents easy to follow guidelines for following the step and concludes with an example. A bibliography is included at the end of the Guidelines for those teachers who would like to know more about the theory from which the five step model was derived.

The Five Step Guidelines to Diagnostic Test Developent

1. List and sequence the skills and tasks that lead up to those you wish to test (the preprequisite skills).

   Section 1. Page 6.

2. Create a description of the skills or content to be tested.

   Section 2. Page 17.

3. Write test items that fit your description in #1 above using the conventions of item writing.

   Section 3. Page 29.

4. Review test items to be sure they fit the test description and conform to item writing conventions.

   Section 4. Page 31.

5. Tryout (field test) items to determine which items need revision, whether the test description needs adjusting, whether the correct sequence of prerequisite skills has been identified, and how many items should be included on the diagnostic test.

   Section 5. Page 36.

STEP 1. SPECIFY THE DIAGNOSTIC SKILLS TO BE TESTED

Importance of the Skill Map

This first step in developing a diagnostic test is often the most difficult. For a test to be useful in planning instruction for individuals

or groups of students, it must assess skills related to the mastery of particular instructional objectives.

It is not always clear just which skills are diagnostic for specific objectives. The skill map will depend upon how test obejctives are defined. For example, when developing a diagnostic reading comprehension test for fourth graders, should the teacher include phonic and structural analysis skills ? How important is the understanding of vocabulary from the content areas to fourth grade reading success? Does reading comprehension include understanding pictures, maps, charts and graphs?

Depending upon the teacher's definition of "comprehension," the skill map may take any number of forms. The teacher who includes comprehension of material in the content areas as well as basal reading and literature in her definition of "reading comprehension" will produce a different set of diagnostic skills than one who defines comprehension as ability to decode and understand basal materials.

How to Develop the Map: Task Analysis

A number of researchers have addressed the problems of developing skill maps and have developed strategies for identifying diagnostic skill continuums. Most of these strategies require teachers to perform a task analysis of the instructional objective for which the diagnostic test is to be created.

Task analysis is a procedure for breaking down an instructional objective into a set of essential subskills leading to the acquisition of that objective. The task analysis is used to create a skill map that

arranges three types of instructional objectives, target, enroute, and prerequisite, into a logical sequence for instruction or testing. A target objective is ucually an important instructional objective that takes a rather significant period of time to teach. Unit and course objectives are generally identified as the target objectives of a diagnostic test. Task analysis is performed upon target objectives which are in turn broken down into enroute and prerequisite objectives.

Enabling objectives describe the subskills students must master in order to demonstrate the performance or knowledge specified by the target objective. It is enabling objectives that provide the basis for instructional units or even specific lessons. Since students must master enabling objectives before they can be be successful on the target objective, these enabling objectives are diagnostic. When enabling objectives are sequenced logically, as from easy to difficult, they form the diagnostic skill map which in turn is the outline for the diagnostic test.

Prerequisite objectives are those we assume students have mastered before beginning our course. All students come with a set of previously learned skills and knowledge. As teachers, we build upon prerequisite skills when planning instruction, be it childrens' ability to say their names and tie their shoes or their ability to write a competent essay. Prerequisite obejctives are an important part of the skill map because they provide the starting point for the diagnostic test. Since students taking the test are assumed to have mastered the course or unit prerequisites, items incorporating these skills may be included on the test. We can

assume the students know something; the prerequisite skills specify what they should already know with reference to a particular set of enabling objectives. The specification of prerequisite objectives is especially helpful when teachers find students are weak in all diagnostic skills tested. The teacher can then assess mastery of prerequisite skills and determine if remediation must include skills not assessed on the diagnostic test.

Task Analysis Strategies

Robert Gagne (1970) identified three kinds of task analysis that could be used to create a diagnostic skill map of course or unit objectives: information processing analysis, task classification, and learning task analysis.

An information processing task analysis, according to Gagne, is a description of the sequence of steps needed to perform a skill, not unlike a computer program. Such an analysis is linear and in many ways doesn't demonstrate the complex relationship between skills needed to perform complex tasks. While it might be relatively simple to describe the steps involved in editing an essay using a word processor (first students insert the software diskette into the disk drive, then they turn on the machine, and so on), the information processing sequence needed to describe how a student revises an essay would be very difficult to specify given the recursive nature of the writing process and the need to specify what evaluative judgments the student makes during the revision process and exactly how those judgments lead to specific kinds of essay revisions.

Samples of information processing task analyses, essentially flowcharts, are available for mathematics (Resnick, 1976; Greeno, 1976),

reading (Resnick & Beck, 1973), and balancing a checkbook (Merrill, 1971).

A classification task analysis is one familiar to teachers. Bloom and Krathwohl's taxonomies of affective, cognitive, and psychomotor objectives are frequently referred to in instructional materials and test manuals. Teachers should be aware, however, that these taxonomies were created to describe test items; they are not a model of how students learn. In fact, studies have shown that cognitive domain levels, recall, comprehension, application, analysis, synthesis, and evaluation, are not necessarily hierarchical. Students may be able to apply a concept without recalling the definition. It may be easier to evaluate an essay than create one (synthesis). Application level test items (e.g., inferential comprehension, word problems) have been found to be easier for some students than recall items (e.g., literal comprehension, recall of concepts, rules and principles or number facts).

A third strategy for breaking down a target objective into diagnostic (enabling) objectives is through a learning task analysis. Unlike a classification task analysis which is useful in describing a skill, the learning task analysis focuses on the sequence of skills. While at first, the learning task analysis may seem indistinguishable from an information processing task analysis which also highlights the skill sequence, it is a more sophisticated approach. Skills in a learning task analysis are sequenced hierarchically not linearally. Each skill in a learning skill map is based upon the one before it and theoretically cannot be mastered until its prerequisite has been learned. The learning prerequisites may include specific cognitive strategies, information, attitudes, and even

motor skills students must possess in order to perform the skill specified in an instructional objective. Because a learning task analysis describes a number of conditions of learning, skill maps generated by this method provide a more thorough diagnosis of student strengths and weaknesses than one created through a classification or information processing analysis. In practice, teachers use a combination of the three to describe the skills needed to attain unit or course objectives.

Specifying the Skill Map

The skill map is created by using one or more of the task analysis approaches described above to specify the prerequisite skills, enabling skills, and performance contexts related to the target objective. The three steps in specifying a skill map are summarized in figure 3.

Describe the testing context. Along with considering which skills to test, the teacher should think about the range of difficulty that the diagnostic test will encompass. Three variables affect task, hence test item, difficulty: the linguistic complexity of the testing material; the cognitive complexity of the task students are asked to perform on the test; the level of discrimination among concepts and topics required to answer test questions; and the level of prompting offered.

Perhaps the most obvious aspect of test context is the lingistic complexity of the items. For example, when diagnosing a student's ability to generalize from reading material, the linguistic complexity of the passage must be taken into account. If all passages used on the test items are quite simple, students may have no trouble with questions requiring them to make generalizations, but still be unable to answer such questions in their social studies or science textbooks. Thus, when diagnosing

FIGURE 3

Step 1. Specify the Skill Map

1.1 Describe the test context
- o Linguistic complexity
- o Cognitive complexity
- o Required level of discrimination among concepts
- o Level of prompting on cues given students about what to do

1.2 Identify prerequisite skills needed to master enabling skills
- o What does a student need to be able to do to attain a given skill?
- o What concepts, rules or procedures must a student know?

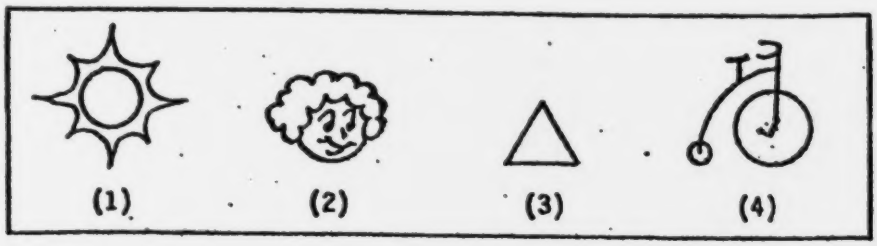
1.3 Write the skill map
- o Skills become objectives to assess on the diagnostic test
- o Identify major points where critical skills are attained
- o Create a test description for each diagnostic point on skill map

reading comprehension, the teacher needs to include passages that are both simply written and some that are linguistically complex. If the skill being diagnosed, such as the ability to solve math word problems, does not rest on the student's ability to decode texts at different levels of linguistic complexity, the teacher will want to keep the language simple so that reading ability does not affect test performance.

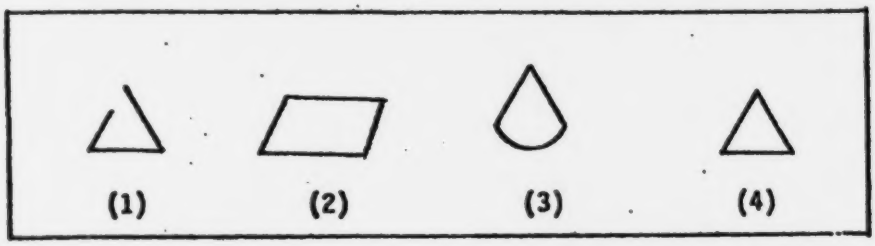
A second aspect of test context is the cognitive complexity associated with demonstrating a skill. Cognitive complexity is the level of "processing" or thinking and the number of cognitive steps a student would need to complete to answer the test question that assesses a skill. When assessing students, the teacher needs to know if students can demonstrate the skill through the full range of cognitive complexity. For example, a diagnostic test of pronoun usage asks students to read a short passage and identify the pronoun that will complete the sentence. This task requires that students demonstrate their knowledge of pronoun usage by using passage context to identify a referent and then to select the correct pronoun for the referent. Students who cannot correctly complete the sentences may still have a working knowledge of pronoun , but they will only be able to demonstrate what they know when the test items are cognitively more simple. A simpler task in this case would be to present students with sentences in which the subjects or objects were underlined and have them select the appropriate replacement pronoun. If the diagnostic test is confined to tasks of the same level of difficulty, the teacher will not know just how much the top students can do nor what the low students already have mastered.

A third context variable is the level of discrimination among concepts or topics that students are required to make. For example, consider the following two test items asking students to identify a triangle (from Baker & Herman, 1983):

A.



B.



Both items test the concept "triangle" but the second required finer discrimination. In the second item, students must be able to apply features of three sidedness, closed figure, and linear figure, while in the first, all a student must know is that the triangle is a linear figure.

Another example of how level of discrimination affects task difficulty appears in the following social studies items:

1. Which country is more democratic, Italy or France?
2. Which country is more democratic, the U.S.A. or U.S.S.R.?

The discrimination required in item one is much finer than that required by item two. Students must test the definition of democracy against more criteria to answer item one correctly than to answer item two correctly.

Closely related to the level of required discrimination is the level of prompting and/or cues given to students about what they are supposed to do on the test. Suppose that a student is given life-like problems to solve using principles of physics but is not told which principles to apply.

Such a test would be more difficult than one in which students are given a principle and told to apply it to solve the problem.

Identify prerequisite skills and knowledge. The previous section identified testing contexts that affect the difficulty level of the skills being assessed. Concurrent with the consideration of factors that may influence skill difficulty is the identification of the skill hierarchy itself using one of the task analysis approaches outlined above. The guiding question in the identification of prerequisite skills or the enabling objectives leading to the target objective being diagnosed is: What does a student need to be able to do in order to perform this skill?

If a teacher wishes to develop a diagnostic test of students' ability to identify the main idea, prerequisite or enabling skills might be identifying details stated in the passage, making inferences from passage details, and understanding vocabulary in context. The prerequisite skills for a test diagnosing students' ability to subtract whole numbers might include the ability to subtract without regrouping, the ability to subtract when regrouping is required in non-adjacent columns and the ability to subtract when regrouping is required in adjacent columns (Gagne, 1977).

Other important questions in the identification of prerequisite skills are:

What rules, procedures, and/or principles must students know in order to attain the target objective?

What concepts must be understood?

Does mastery of the target objective require knowledge of certain facts?

Each of these questions identifies potential diagnostic points in the skill map of the target objective. These diagnostic points are crucial because each becomes the basis for a set of items assessing mastery of an enabling/prerequisite skill.

Write the skill map. The enabling/prerequisite skills chosen through task analysis should be recorded on a skill map. When the teacher does not have the time to do a detailed task analysis, the skill map may actually be the scope and sequence for a course or the continuum provided in the reading, mathematics, or language text. District courses of study often provide skill maps that can guide the development of diagnostic tests. It is important to realize, however, that these materials were developed through one or more of the task analysis procedures described in this manual.

The most difficult step in developing the skill map is the identification of major points where critical skills are obtained, for each of these points is the basis for a test description from which diagnostic items will be written. How many enroute skills should be specified on the map? On the one hand, the more that are included, the greater the diagnostic potential of the test. On the other hand, each diagnostic point adds greatly to test development and administration time as well as test length. It may be faster to carefully reteach a skill specified at a more general level to an entire group than to create an extremely fine grained and lengthy test that will uncover the problems of each student in the class. Two or three enroute skills probably provide enough diagnostic information for students without serious learning handicaps.

Using the Skill Map to Develop the Test

Once diagnostic assessment points have been specified on the skill map, the teacher may begin to write test descriptions for each of these points. The test description identifies the item types, item content and item writing rules for each enroute skill appearing on the diagnostic test. The procedure for developing a thorough description is the subject of Step 2 in this manual.

STEP TWO: CREATE THE TEST DESCRIPTION

This second step of the test development process is the most most creative and often the most arduous. Developing a test description requires hard thought about the nature of the skills, item content, and item formats that will appear on the test. Because test development and administration requires much effort, the skills that will appear on the diagnostic test should represent major diagnostic points on the skill map.

Why A Test Decription

The test description standardizes the test items and makes the item writing process easier. Descriptions provide detailed guidelines for item content, level of difficulty, and answer choices (for multiple-choice items) or how answers are to be scored (written answers). The test description states the test objective clearly enough for other teachers to understand exactly what skills the test is assessing. It also includes a sample item so that others can visualize what the test will look like. In short, the description provides instructions about how to write the test.

A test description is written for each enabling objective identified as a crucial diagnostic point on the skill map. The following procedures

>or writing test descriptions will be repeated for all objectives appearing on the diagnostic test. Sample test descriptions appear in Appendix 1.

2.1 Identify Objectives Worth Testing

A test description is written for diagnostic skill that will appear on the test. Thus, objectives selected for the test should be restricted to those for which the process of writing a complete description will be worthwhile.

Most school districts have curriculum guides or written scope and sequences for required courses. Often adopted texts and other materials include objectives and suggest an appropriate instructional sequence. Test objectives will generally be selected from these sources. However, course guides and commercially developed instructional materials include so many objectives that culling out the most important ones to test is no small task. The following criteria identify objectives worth testing.

Time: How much instructional time does it take to teach the objective? Select objectives that cover a reasonable amount of instructional time: a chapter, a unit, a term.

Higher Order Skills: How does the objective relate to higher order skills? Recent reports on the status of American education have criticized the level at which some instruction occurs. Be sure that your objectives encompass or are prerequisite to higher level thinking, problem solving and other important educational goals.

Long Term: How does the objective relate to long-term curricular goals? Be sure objectives are part of a coherent strand of learning.

Importance: What is the intrinsic importance of the objective? Is it a key skill or a trivial learning task?

2.2 Specify The Skill Required To Meet The Objective

Curriculum objectives don't always specify how a student will demonstrate mastery of the skill or content. Typically objectives resemble the following: The student will draw conclusions about written passages. What's missing is the specification of what the student is to do. How does one know that the student is drawing a warranted conclusion?

Clarification of this objective involves specifying some task, such as: Students will read a passage and compose written answers to four "why" questions that require conclusions be drawn from the passage.

As objectives have various levels of importance and cognitive complexity, so do the skills required to meet the objectives. Several taxonomies have been written to help test developers classify the skill level of objectives, the most familiar of which is Bloom's Taxonomy (1956). According to Bloom, skills may be classified into six categories: recall, comprehension, application, analysis, synthesis or evaluation. Higher order skills include application, analysis, synthesis and evaluation.

Gagne (1970) presents a diffent method of identifying which skills are higher order. His taxonomy is based on the student's ability to demonstrate knowledge of concepts, rules, procedures and the application of these to solving problems. Application of rules or procedures and problem solving are the highest level of skill in Gagne's scheme.

Many other skill taxonomies exist which may be useful to test developers. The range of identification of the levels of cognitive development (Piaget) to levels of moral development (Kohlberg) both of

which are beyond the scope of these guidelines. What is important to remember is that there are several approaches for identifying and specifying skills that will appear on a diagnostic test, just as there are several approaches for creating the skill map by breaking down the target objective into enabling and prerequisite skills.

2.3 Specify The Content To Be Covered

As sources of test objectives are reviewed, the actual topics or content to be tested must also be considered. For most teachers, curriculum materials and adopted texts will dictate test content. The following questions suggest guidelines for selecting content from these materials.

Context: In how many different contexts will students need to apply the skill? For example, if the skill to be tested is reading comprehension, will the passage selected be narrative or expository? Will the main idea in the passage be implicit or explicit? If the test content involves problem solving as in math or physics, will students need to apply the prinicples in laboratory settings or in life-like settings? If the skill specified is analysis of historical events, how many and what types of events will students be required to analyze?

Scope of Information: What information will students need to know? Is there a list of concepts, vocabulary or facts that students need to acquire? For example, if the skill involves identifying parts of the human body, how many and at which level of generality are acceptable? If the skill requires students to analyze the major events of the Civil War, what types of events are considered major?

Quantity of Information: How many different topics will be used to test student skills? Will the topics be ones with which students are familiar or will students be presented with new information (this is especially likely when testing problem solving skills)? Is the purpose of the test to assess student skill or knowledge of information? The amount of information included on the test will differ according to each purpose.

2.4 Select Appropriate Item Type

Once the objective is clarified and test content delimited, item types may be considered. There are no hard and fast rules for choosing the appropriate item format for a specific skill, although there is sometimes an inverse relationship between the ease with which an item is constructed and scored and its measurement validity. For example, students have a 50% chance of guessing the correct answer to easily scored true-false items. On the other hand, essay tests provide the most valid measure of student writing skill and divergent thinking and are quite difficult to score.

There are three broad categories of test items, selected response, constructed response, and performance. Selected response items present students with questions and a set of alternatives. Students select the answer from those given. Examples of selected response items include true/false, multiple choice and matching. Constructed response items present questions or problem and require students to write out their own answers or solutions. Essay tests, math tests requiring students to derive their own answers, short answer, completion, and close tests are examples of constructed response items.

The third category of test items, performance, is rarely encountered in standardized testing programs. A performance item presents the student with a task, either life or simulated, and asks the student to complete the task while the instructor rates techniques used as well as the solution arrived at. Physical fitness tests, laboratory experiments, drafting problems, cooking, sewing, or auto mechanics tasks, as well as diving, dancing, speech, drama, and spelling competitions are a few of the many ways students are asked to show competence by performing.

Selecting an appropriate item type is a two step process involving both creativity and a thorough knowledge of the skill being tested. The first step is to brainstorm as many ways as possible to test a certain skill. Next the most appropriate method is selected balancing concerns about valid assessment against the time available for testing. When selecting item formats, the testing method, paper-pencil, computer, performance, or oral response should be considered so that the item will conform to existing constraints. For example, items administered by computer must be written to fit the computer screen, either 40 or 80 columns long and no longer than 25 lines of text.

The appendix includes examples of different item formats appropriate for multiple choice testing in reading, math, science and social studies.

Once an item format has been selected, a few illustrative items and directions to students should be written. The item and directions provide the model for all subsequent test items assessing each test objective.

2.5 Write the Test Description

Once the sample items have been formulated, the test description can

be written. The format for the test description should use terminology and be organized in a way that communicates with teachers. Different researchers have suggested slightly different formats. The one appearing in figure 4 and described in this section combines models suggested by Popham (1980) and Baker (1974).

The General Description and Sample Item simply state in writing the test objective and the sample item with directions to students. The Content Limits codify the decisions made in Step 3 of the Test Description process, "Clarify Content to Be Covered." Content Limits include rules for writing test questions, rules for selecting the topics or information to appear in test questions, and a description of the kinds of reading materials, pictures, charts, graphs or other prompts that will be presented to students taking the test.

Content Limits for multiple choice items define the eligible content and format of the item. The rules for item content are particularly important. By describing the entire range of content to be tested, the teacher will be able to determine which concepts are being only partially learned and which are mastered.

Content Limits for written responses describe the questions, topics or writing prompt presented to students, instructions to students about how to answer, and the test setting (when appropriate).

Response Limits provide rules for writing the student answer choices for multiple-choice items or judging answers for written tests. For

Figure 4

Sample Test Description

General Description: When presented with a short paragraph from which names or referent pronouns have been deleted, students will select from a list of four alternatives the correct pronoun.

Sample Item

Mary and Paul jumped up and cheered the touchdown. _____ were excited that the team was winning.

a. we
b. he
c. they
d. them

Content Limits:

1. The student will be presented with a short (3-5 sentences) paragraph describing an action or event involving two or more named individuals.

2. A blank will replace the named individual in one sentence.

3. Students will be presented with the question: Which pronoun correctly completes the sentence?

4. Paragraphs and deletions will be written to test the following rules:

   a. When the pronoun is the subject of the sentence or clause, it should be in the nominative case.

   b. When the pronoun is the direct object, it should be in the objective case.

   c. When the pronoun is the indirect object, it should be in the objective case.

(The rules for item content exemplified in #4 above are particularly important. By including paragraphs to test each pronoun rule, the teacher will be able to determine which are causing students difficulty and can plan instruction based on this information.)

Response Limits

1. Five alternatives will follow each paragraph: The correct answer and four incorrect choices.

2. The correct answer will be an example of the correct application of one of the pronoun rules listed in the content limits.

3. Incorrect answers (distractors) will be written to exemplify each of the following kinds of wrong answer choices:

   a. a pronoun in the correct case but incorrect number/gender.

   b. a pronoun in the incorrect case, but correct in number and gender.

   c. a pronoun representing an incorrect referent but correct in case, number and gender.

   d. a pronoun in the incorrect case, incorrect in number and/or gender.

Note that each choice represents a potential error. When clusters of students choose "a", instruction can be targeted to agreement of number and gender. Specific and systematically written wrong answer choices provide rich diagnostic information.

multiple choice or selected response items, the rules for constructing the correct answer and wrong answer choices appear in the response limits. In order for the item to provide diagnostic information, wrong answer choices should represent common student errors. Specific and systematically written wrong answer choices provide rich diagnostic information.

Response limits for essay or constructed reponse items list the criteria for judging an answer correct. Criteria may be holistic such as the rubric for scoring a first hand biography essay displayed in figure 5 or more specific content outlines or analytic scoring scales. Although creating response limits for essay, short answer and other constructed response items can be time consuming, the discussion involved in developing scoring guides clarifies both the skill being tested and the kinds of material that should be covered when instructing to the objective.

Test Descriptions take time and effort to write. However, when finished, the teacher has a guidelines for developing test items and for planning instruction. The Test Description can be used to plan classroom activites and appropriate practice on important skills. The Test Description is simply an elaborated lesson plan, one that a teacher might create for a substitute who knew nothing about the subject and who was responsible for an entire unit of instruction.

The procedures for creating a test description are summarized in figure 6. A review form for checking the Test Description is reproduced in Appendix 1. This form will be useful when developing a Test Description for the first time. Review questions will prompt teachers for information that should be included in their Descriptions.

Figure 5

Grade 10

Rubric for Fall Semester Composition - Firsthand Biography

Task: Student writes a multi-paragraph essay which explores the impact of a significant change or conflict in the life of either a personal acquaintance or a fictional character. The paper should be chiefly an examination of the incident's causes and the changes it brought about in the life of the character; it should not simply be a narrative telling of the pivotal incident. However, the details of the central incident or conflict should become clear in the course of the essay.

5: Applies to papers that do everything well. Includes all criteria of a 3 paper and, in addition,

a) is informative and entertaining
b) is insightful
c) has an engaging beginning and an effective conclusion
d) effectively conveys the uniqueness and power of the experience in the life of the subject
e) uses language well: vivid, precise word choice; sophisticated vocabulary; varied sentence structure
f) effectively organizes ideas in several coherent paragraphs; employs transitions well
g) exhibits excellent proofreading/editing skills.

4: Meets all of the criteria of a 3 paper, and has one or two characteristics of a 5 paper.

3: Meets all of the following criteria:

a) Content:
- focuses on the causes of the selected incident and the changes it brought about in the life of the subject (is not simply a narrative of an incident)
- uses predominantly appropriate word choice

b) Organization:
- uses paragraphs to organize material logically
- has a distinct beginning, middle, and end
- has a title

c) Mechanics:
- is generally free of mechanical/grammatical errors
- has neat appearance: white, lined paper, blue/black ink, legible cursive/manuscript writing using upper and lower case letters
- has at least 300 words
- has at least three paragraphs.

2: Meets some but not all of the criteria of a 3 paper

or

contains mechanical/grammatical errors which interfere with the reading of the paper.

1: Fails to address the task as directed

or

is illegible.

Revised 7/83

FIGURE 6

Step 2. Create the Test Description

2.1 Identify objectives worth testing
- o Amount of time: unit, semester, year
- o Higher order skills
- o Long term curricular goals
- o Intrinsic importance

2.2 Specify skills required to meet objectives tested
- o What do students have to do?
- o What is the likely sequence?
- o What is the level of cognitive complexity?

2.3 Clarify content to be covered on test
- o Contexts in which the skill will be demonstrated
- o Scope of information to be included
- o Quantity of information to be included

2.4 Select item type
- o Brainstorm all possible ways to test
- o Select the most valid type for objective
- o Write sample items and directions

2.5 Write test description
- o General Description: restates objectives and skill levels
- o Sample Items and Directions
- o Content Limits: specifies context, scope, quantity of information included in item and provides guidelines for writing item stem
- o Response Limits: Specifies rules for writing correct answer and wrong answer distractors and/or rules for scoring written responses.

STEP 3. WRITE THE TEST ITEMS

Once test descriptions have been completed for each important enabling skill selected from the skill map, developing the actual items is simply a matter of following the directions given by the descriptions. Teachers may find that as they write items or incorporate course content into the test, they are adding content or developing response choices that differ from those specified in the test description. When this occurs, the description should be modified to include a description of the new content. Test descriptions provide guidelines, not mandates, for item development. They, like the items themselves, must be tested for their ability to describe all of the contingencies of the testing situation. When there are gaps in the descriptions, simply revise them to fit the new item requirements.

An important decision is the number of items that should be written to match any one test description, as well as the number of items that should appear on the diagnostic test itself which encompasses several skills. How many items should be written? As many as possible. Some testing companies develop two or three items for each one that appears on the finished version of a test. Since diagnostic tests require more items per skill than tests reporting only summary or survey information, the need for a fairly large pool of items is further supported. A rule of thumb suggests that anywhere from three to eight good items per skill may be required for diagnostic purposes. Since as many as one half to two thirds of the items written may have some kind of editorial or statistical flaw, for each test description as many as 24 items may have to be drafted in order to select eight good ones.

FIGURE 7

Step 3. Write Test Items

3.1 Follow rules in Test Description to draft items

3.2 Review items against Guidelines for Item Writing to avoid technical flaws

3.3 Write three to five items for each subskill, content and context specified on the Test Description

Once items are drafted, they should be reviewed against the item writing rules in the test descriptions for consistency. When a mismatch between item and description occurs, the description may be changed or the item thrown out.

Additionally, items should be reviewed against the item quality guidelines developed by Gronlund (1968) and O'Neil (1979). Typical item quality guidelines are summarized in figure 8. These guidelines help teachers to flag items that may enable a student to give the correct answer without really knowing how to respond correctly. The guidelines also insure that ambiguous or tricky language is edited out of the items so that students who may have mastered the skill aren't misled into giving an incorrect answer.

Although the first review of the test items is done by the item writer, it is hardly sufficient to catch all problems that may occur either with matching the item to the test description or with poorly worded questions. The next step is a formal item review.

STEP 4. REVIEW TEST ITEMS

The formal test item review should be done by persons other than the item writers. The purpose of the review is to screen out poorly written questions and ones that don't assess the skill as intended. Two review questions suffice to identify these two kinds of poor items (figure 9):

- Do the items match their test descriptions?
- Are they free from technical flaws; do they follow conventional rules for item construction?

Figure 8
General Guidelines for Item Writing

Typical Rules for Multiple Choice Items:

1. The stem of the items should be meaningful by itself and should present a clear problem.

2. The stem should be free from irrelevant material.

3. The stem should include as much of the item as possible except where an inclusion would clue the responses. Repetitive phrases should be included in the stem rather than being restated in each alternative.

4. All alternatives should be grammatically consistent with the item stem and of similar length, so as not to provide a clue to the answer.

5. An item should include only one correct or clearly best answer.

6. Items used to measure understanding should contain some novelty and not merely repeat verbatim materials or problems presented in instruction.

7. All distractors should be plausible and related to the body of knowledge and learning experiences measured.

8. Verbal associations between the stem and correct answer or stereotyped phrases should be avoided.

9. The correct answer should appear in each of the alternative positions with approximately equal frequency and in random order.

10. Special alternatives such as "none," "all of the above" should be used sparingly.

11. Avoid items that contain inclusive terms (e.g., "never," "always," "all") in the wrong answer.

12. Negatively stated item stems should be used sparingly.

13. Avoid alternatives that are oppositive in meaning or that are paraphrases of each other.

14. Avoid items which ask for opinions.

15. Avoid items that contain irrelevant sources of difficulty, such as vocabulary, sentence structure.

16. Avoid interlocking items, items whose answers clue responses to subsequent items.

17. Don't use multiple choice items where other item formats are more appropriate.

Figure 8 (continued)

Typical Rules for Short Answer and Completion Items:

1. A direct question is generally better than an incomplete statement.
2. Word the item so that the required answer is both brief and unambiguous.
3. Where an answer is to be expressed in numerical units, indicate the type of units wanted.
4. Blanks for answers should be equal in length. Scoring is facilitated if the blanks are provided in a column to the right of the question.
5. No grammatical cues should be give, e.g. a _____; an _____.
6. Where completion items are used, do not leave too many blanks.
7. For completion items, only key words should be left blank. Leave blank only those things that are important to remember.
8. In composing items, don't take statements verbatim from students' textbook or instruction.
9. The scoring key should anticipate possible synonyms or acceptable variants at the desired response.

Typical Rules for True-False or Alternative Response Items:

1. Avoid broad general statements for true-false items.
2. Avoid trivial statements.
3. Avoid negative statements and especially double negatives.
4. Avoid long complex sentences.
5. Avoid including two ideas in a single statement unless cause-effect relationships are being measured.
6. Avoid questions which include indefinite terms, degrees or amounts.
7. Include opinion statements only if they are attibuted to particular sources.
8. True statements and false statements should be approximately the same length.
9. The number of true statements and of false statements should be approximately equal.
10. Avoid taking statements verbatim from students' text or instruction.
11. An item's truth or falsity should not depend on an insignificant word or phrase.

FIGURE 9

Step 4. Review Test Items

4.1 Do items match specifications on Test Description?
- o Have colleagues review items with Test Description in hand

4.2 Are items free of technical flaws?
- o Check to see items don't violate general item writing rules
- o Check accuracy of content tested
- o Have several colleagues take items to check correctness of answer keys and defensibility of correct response

Do the Items Match Their Specifications

The answer to this question establishes in part the content validity of the diagnostic test. Each item should be examined by a colleague to compare its adherence to the item guidelines provided by the test descriptions.

The item format should conform to that described and illustrated by the sample item. Content in the item stem (question, writing topic, problem, etc.) should conform to that described in the content limits. Correct and incorrect answer choices (for selected response items) and scoring guidelines (for constructed response items) should conform to the rules set down in the response limits section of the test description.

Although the technical edit is designed to catch language that is ambiguous or tricky, such as the use of double negatives, unnecessarily difficult language should be flagged as well.

It is especially important that reviews try to follow item directions in order to check their clarity. Whenever possible, directions should be simple and item formats fairly self explanatory.

The test description may be changed at the formal item review stage as well as the item writing stage. Items also may be revised after the formal review in order that as many good items as possible be available for field testing.

Are the Items Free of Technical Flaws?

The second stage of the item review checks to see that items don't violate conventional item construction rules. Items may be revised at this point, but if several items have the same problem, such as unplausible distractors, the fault may lie with the test description rather than with

how the item was written. Thus, the technical edit may indicate a revision of the test description as well as the item.

Item review is the next-to-last step in the test construction process. All that is left to do is to tryout the revised items with groups of students to see how the test questions actually behave.

STEP 5. TRYOUT ITEMS

The item tryout, sometimes called a pilot or field test depending upon the number of students involved, is the final step in the test development process. Optimal tryout procedures include the three steps presented in figure 10. First items are tried out with a small group of students who are representative of the students with whom the test will be used. The tryout sample should be small, from five to twenty students, and need not be randomly selected. Students will take pilot versions of the diagnostic test so that items can be tried out "in context."

This initial tryout should be used to gather as much information as possible about how the test is functioning. Teachers should note the problems students have following test directions, their frustration level during the test, and the clarifying questions they ask. Students are especially good at pointing out items that have either no or many correct answers. This feedback will be used to revise, yet again, faulty questions or constructed response scoring procedures.

The most important statistical information from this pilot test is the calculation of item difficulty values, the proportion of students answering an item correctly. Item difficulty indices can range from 0-1.0. When multiplied by 100, the difficulty indices become the percent of students answering an item correctly. The difficulty values can be used to

FIGURE 10

Step 5. Tryout Items

5.1 Choose a small sample of students (5-20) to take items
- o Select appropriate level and age of students for sample
- o Check clarity of directions
- o Check answer keys
- o Note problems students are having with test by questions they ask

5.2 Revise items and directions as indicated by tryout

5.3 Field test items with large sample (at least 100)
- o Select appropriate level and age for field test
- o Try to get sample representative of range of ability background and ethnicity of actual students with whom the test is to be used
- o Have other teachers administer test using "standardized" directions and time limits
- o Look at patterns of student answers to analyze items for
  1. Difficulty
  2. Sequencing from easy to difficult
  3. Homogeneity (are items measuring the same skill (acting alike)

tentatively identify "hard" and "easy" items. Item difficulties provide a rough check on the task analysis used to develop the diagnostic test. More students should answer items correctly which measure prerequisite or less complex skills than items assessing complex behaviors and knowledge. When the item statistics call into question the logic upon which the skill map was constructed, look first for a poorly written item and then resassess the task analysis.

Item difficulties are helpful in detecting items that are not diagnostic. If the items measuring a particular enabling skill are easy (80% or more of the students answered them correctly), then there is little to diagnose.

A second statistic to check at the pilot stage is the proportion of students selecting each answer choice for selected response items. Distractors that no one selects don't provide diagnostic information. One way to identify wrong answer choices that are likely to be selected by students is to present students with the multiple choice questions without the answer choices and have them supply an answer. The student-generated answers can then be categorized and paraphrased to represent the three or four most common student errors.

Once items have been tested and revised with a pilot group, a large scale tryout should be planned. This tryout, often called the field test, is a more formal undertaking and should follow procedures used in standardized testing programs. Formal directions to teachers should be reproduced and test administration conditions kept as similar as possible among teachers participating in the field test. A least 100 students should be selected to take the diagnostic test and these should be representative of

the intended population. Whenever possible students participating in the tryout should have similar instructional histories, that is they are exposed to the same curriculum. When the test developer cannot be sure that the students participating in the field test represent a similar stage of skill development, random selection or the use of many classrooms with the classroom being the unit of data analysis, should be used.

Item analyses conducted after the field test are more formal than those following a pilot tryout. At this stage, items are reviewed for "homogeneity," how much items designed to measure the same skill resemble each other statistically, difficulty, and their contribution to accurate diagnostic decision making. A description of the analytic procedures used to assess the above item characteristics is beyond the scope of this manual. However, the large number of students involved in the field test and the complexity of the calculations required to evaluate items statistically suggests that teachers must have access to professional test scoring services, a large computer, or a microcomputer attached to a scanner and equipped with special item analysis programs in order to analyze field test data.

REPRISE

Teachers use diagnostic tests to identify student needs for remediation or acceleration, to plan instruction for individuals and groups of students and to select effective instructional materials and strategies. Because of the costs attached to diagnostic test development and their specificity to particular curricular objectives, published diagnostic tests are nearly impossible to find.

This manual describes a five step procedure which teachers can follow to develop their own diagnostic instruments to be used with either individuals or groups of students. Examples used in this manual were drawn from the cognitive domain, but the procedures are equally applicable to the development of affective or psychomotor tests.

Much space in this manual is devoted to the conceptual plan underlying the diagnostic test, the skill map. Several strategies for developing the skill map through task analysis are mentioned. At the heart of diagnostic testing technology is the application of research-based analytic methods to the description important skills. Description, indeed, is the the core of the process and the procedure that separates diagnostic test development strategies from those used in survey and competency testing programs. Thus, before even one item is written, teachers are asked to spend time and thought to descriptive information; first, the diagnostic skill map and later the test description that specifies how each skill will be measured.

Unlike nationally normed tests which have a test description for each subtest or content area, the test descriptions used to generate diagnostic items specify one item type and skill only.

Diagnostic tests, even though they focus on skills even more specific than those covered by test publishers, require a set of test descriptions. Once the conceptual base for the diagnostic test is documented in the skill map and the test description, the steps for writing, reviewing and trying out diagnostic items remain much the same as those used in survey testing programs.

Even when teachers are unable to construct their own tests using the method presented in this manual, a knowledge of the procedure will help them evaluate the quality of tests developed by their colleagues or publishers.

Bibliography

Baker, E.L. Beyond objectives: Domain-referenced tests for evaluation and instructional improvement. In Hively, W., Domain-referenced testing. Englewood Cliffs, New Jersey: Educational Technology Publications, 1974.

Baker, E.L. & Herman, J.L. Task structure design: Beyond linkage. Journal of Educational Measurement, V. 20, N. 2, Summer 1983.

Bloom, B.S., Englehart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (Eds.) Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay, 1956.

Conoley, J. & O'Neil, H. A prime for developing test items. In O'Neil, H. (Ed.), Procedures for instructional systems development. New York: Academic Press, 1979.

Gagne, R.M. Analyses of lectures. In L. Briggs (Ed.) Instructional design: Principles and applications. Englewood Cliffs, New Jersey: Educational Technical Publications, 1977.

Gagne, R.M. The conditions of learning. (2nd edition) New York: Holt, Rinehart, and Winston, 1970.

Gagne, R., & Biggs, L. Principles of instructional design. Holt, Rinehart and Winston: New York, 1979.

Greeno, T.G. Cognitive objectives of instruction. Theory of knowledge for solving problems and answering questions. In O. Klar (ed.), Cognition and Instruction. Hillsdale, NY: Erlbaum Associates, 1976.

Gronlund, N.E. Constructing achievement tests. Englewood Cliffs, New Jersey: Prentice Hall, 1968.

Merrill, P.F. Task Analysis in information processing approach: Technical items #27. Tallahassee, Florida State University, CAI Center, 1971.

Popham, W.J. Domain specification strategies. In Berk, R.A. (Ed.), Criterion-referenced measurement: The state of the art. Baltimore: Johns Hopkins University Press, 1980.

Resnick, L.B. Task Analysis in instructional design: Some cases from mathematics. In D. Klahn (Ed.) Cognition and Instruction. Hillsdale, NY: Erlbaum Associates, 1976.

Resnick, L.B., & Beck, I.L. Designing instruction in reading: Interaction of Theory and Practice. In J.T. Gutherie (Ed.) Aspects of Reading Acquisition. Baltimore: The Johns Hopkins Press, 1973.

APPENDIX 1

- Model Test Description: Science
- Model Test Description: Punctuation
- Model Test Description: Social Studies
- Test Description Review Form

Test Guidelines

Curriculum Objective: E 07 Classify, in a progression, the characteristics of energy types.

Test Objective: Arrange sets of 3 objects producing or exhibiting energy, (heat, sound light-in the form of color or sheen) in a progression from least to most intensity.

General Description: Given a set of 3 objects producing varying intensities of one energy form (heat, sound, or light as color or sheen) students will arrange the objects in progression from the least to the most energy.

Sample Item:

Directions: (To the teacher) With the student, manipulate three objects which produce various intensities of one energy form. While arranging the objects from cool to hot (light-colored to dark; quiet to loud) explain each placement with one sentence, ie.: "This is the coolest (lightest; quietest). It is first. This is medium. It is next. This is hottest (darkest; loudest). It is last."

Present three new objects which produce a different form of energy from your sample. Instruct the student, "You arrange them-first, which? Second, which? Third, which?" Do not give students feedback on their choices. Indicate response on the checksheet. Check=correct. O=wrong. (See Response Criteria.)

*Be aware of children's comfort levels, especially with powerful hearing aides. Adjust students' aides accordingly. Students with very little or no hearing should be encouraged to feel the objects' vibrations. If that is not feasible, such students may be excused from testing for sound.

Sample: The student is given a cup of hot chocolate, a cup of warm water, and a cup of cold water to arrange.

A. Cold water, warm water, hot chocolate

(See checksheet, Response Criteria.)

Test Item Limits: Sets of 3 objects/instruments which produce clearly noticably varying degrees of one energy form (heat, sound light-as color or sheen). For testing heat and light, the objects should be similar in other attributes (i.e., Sample Item; 3 cups of liquids differing in temperature) to prevent the child's confusing the attribute being tested. This may not always be possible for sound.

High-intensity energy forms MUST NOT exceed students' comfort limits.

Students are not to be prompted, with the exception of those similar to, "You arrange these. Which is first?....Which is second?....Which is third? (See Sample Item, Directions.)

Items tested for heat should be tested as soon as possible after being removed f rom the heat source, but students should not see where object was stored. Objects may be wrapped in cloth or paper towels to help them retain heat longer.

See Stimulus Domain Supplement for suggested groups of objects.

Response Criteria: The student will arrange the three objects for each test item in this order: low intensity, medium intensity, high intensity. A correct response will be indicated with a check; an incorrect response with O.

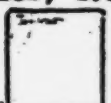
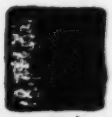
	Trial 1	Trial 2
Heat		
Light-Int.		
Light-Sheen		
Sound		

Test Construction Guidelines:

The test will consist of 8 items; 2 for each: color, sheen, heat, sound intensity. These categories can be tested in any order.

Standard of Performance: The student will correctly seriate 6 out of 8 test items, with at least one item correct for light (either color or sheen), heat and sound.

Stimulus Domain Supplement - E 07

Energy Form	Low Int.	Med Int.	High Int.
Light (color)	Colored shape on paper (triangle, square, circle rectangle, oval) of 1 basic color (red yellow, blue, green, orange, violet) in 3 degrees of intensity (light, medium, dark). Each set should be of one shape and one color, i.e.		
	 (light red)	(med. red)	 (dark red)
(sheen)	Colored shape on paper as described in color, above, but painted with the same intesnity of color in different kinds of paint, i.e.:		
	(flat red)	(glossy red)	(metallic red)
	or small objects of the same color intensity and size but with flat, glossy, and metallic surfaces.		
Heat	Cups, glasses, similar containers of liquid of differing temperatures (styrofoam, ceramic, thermos or insulated containers being the best) i.e.:		
	Styrofoam cup of cold milk water bottle of cold water	Styrofoam cups of warm water water bottle of warm water	Styrofoam cup of hot coffee Water bottle of hot water
	Similar objects (bricks, racks, cups, marbles, dishes, other non-combustibles) that have been in the refrigerator, room temperature, a hot sunny spot, in the oven (@150°) i.e.:		
	Cup of pennies from the refrigerator	Cup of pennies from a sunny spot	Cup of pennies from the oven
	Heating pad off	Heating pad on "medium"	Heating pad on "high"
Sound (intensity)	Bells of 3 different intensities, but of sizes which vary only slightly, so students are not able to arrange by size to correctly seriate by loudness, i.e.:		
	Plastic bell	Aluminum bell	Brass bell
	Drums fitting the above description (or percussive items such as boxes, tins, pans), i.e.:		

Plastic waste-basket	Wooden box	Tin pan
Party noise makers (horns, whistles, clackers, etc.)		
Plastic whistle	Party whistle with rolled paper end	Metal coach's whistle

Grade Level: Grade 9

Subject: English-punctuation

Domain Description: Correctly punctuating given paragraphs adapted from a standard eighth grade text of a practical/informative nature.

Content Limits: The student will be presented with one paragraph in which all the correct punctuation marks have been omitted, except for apostrophes in contractions (I'll), and possessives (Jane's), dashes and semi-colons.

For each question, students will be asked to choose all the correct punctuation marks which must be added in a given sentence to make the sentence correct. The punctuation marks to be identified and added may include:

a. periods at the end of a declarative or imperative sentence, after an abbreviation, or an initial
b. question marks following an interrogative sentence
c. exclamation point after exclamatory sentences or interjections
d. colon after the salutation in a business letter, or to separate minutes and hours in expressions of time, and to show that a series of things or events follows
e. quotation marks enclosing a quotation or a fragment of it, enclosing the title of a story of poem which is part of a larger book
f. comma in a date or address; to set off such words as "yes" at the beginning of a sentence; to set off names of persons or words (phrases) in apposition; to separate words in a series, direct quotations, parallel adjectives, parenthetical phrases; after introductory prepositional phrases; before coordinate conjunctions; and after the salutation and closing in a friendly letter; to separate a dependent clause from an independent clause in a complex sentence.

Distractor Domains: The alternate responses to the questions may include:

a. omission of punctuation mark (s) within a given sentence which should be included, or
b. inclusion of a punctuation mark or marks which is not necessary or correct in the given sentence

Directions: The directions will be given: "Choose the letter which contains all the necessary punctuation marks in the given sentence which will make the sentence correct."

Format: Each question will be multiple choice, with four possible responses.

Sample Item: 1. If she starts to sing again I'll crack up 2. It is funny how it hurts to hold back a laugh 3. I was sitting in the auditorium at 10:00 am and we were having a singing rehearsal for graduation 4. Sit up Get off those shoulders Think tall Sing tall Sing like this said Ms Small 5. I knew that if she was going to tweet like a bird again I would laugh 6. But I just could not laugh because Ms Small would kick me out of the auditorium and that meant Felson's office--and no graduation 7. La la la--sing children Sing with your hearts said Ms Small 8. I couldn't hold it 9. She was so funny I almost rolled of the auditorium seat 10. The other students didn't laugh but me I sounded like Santa Claus 11. It became quiet for a second 12. What are you doing Joe I know it is you Present yourself to Mr Felson at once that voice said 13. Ms Small is a foot shorter than a tall Coke but she has the bark of a hungry hound dog

1. The first sentence should be written:

   a. If she starts to sing again I'll crack up.
   b. If she, starts to sing again, I'll crack up
✓ c. If she starts to sing again, I'll crack up.
   d. If she starts, to sing again, I'll crack up.

SOCIAL INTERACTION
Grades 7-9
INTELLECTUAL PROCESSES

Applying Concepts of United States Foreign Policy

General Description

Given a description of a fictitional international situation in which the United States may wish to act and the name of an American foreign policy document or pronouncement, the students will select from a list of alternatives the course of action that would most likely follow from the given document or pronouncement.

Sample Item

Directions: Read each fictitious example below. Decide what action the United States would most likely take based on the given foreign policy document. Write the letter of the action on your answer sheet.

Some Russian agents have become members of the Christian Democratic Party in Chile. The party attacked the President's house and arrested him. The Russian agents set themselves up as President and Vice-President of Chile. Chile then asked to become an "affiliated republic" of the U.S.S.R.

Based on the Monroe Doctrine, what will the U.S. do?

a. ignore the new status of Chile.
b. warn Russia that its influence is to be withdrawn from Chile.
c. refuse to recognize the new government of Chile because it came to power illegally.
d. send arms to all groups in the country that swear to oppose communism.

Stimulus Attributes

1. The fictitional passage will consist of 50 words or less followed by the name of a foreign policy pronouncement or document inserted into the question, "Based on the _____, what will the U.S. do?

2. The policy named in the stimulus passage will be a document or pronouncement selected from the Domain Supplement.

BEST COPY AVAILABLE

2. Response alternatives consist of the correct response and three distractors. Each alternative will have the following characteristics:

   a. Describe a specific course of action that refers to the people, nations, and actions in the stimulus passage.

   b. Be brief phrases written to complete the understood subject, "The United States will. . ."

3. Distractors will be written to meet these additional criteria:

   a. Each distractor will describe an action derived from a different document or pronouncement selected from the Domain Supplement.

   b. Documents or pronouncements from which identical courses of action may be derived will not be used.

   c. The decision not to act may be used as a course of action when it is based on a document or pronouncement.

4. The correct response will be the course of action that is governed by the principles described in the document or pronouncement named in the stimulus passage.

BEST COPY AVAILABLE

Domain Supplement

Foreign Policy Documents and Pronouncements

The following list of foreign policy pronouncements and documents was selected from Brockway, Thomas, Basic Documents in United States Foreign Policy. Princeton, New Jersey, D. Van Nostrand Company, 1968. The items selected were chosen on the basis of their generalizability and potential application to current events. The list appears in chronological order.

1. The Declaration of Independence
2. Washington's Farewell Address
3. The Monroe Doctrine
4. Webster on Revolutions Abroad
5. Open Door in China
6. The Platt Amendment
7. Roosevelt Corollary of the Monroe Doctrine
8. The Fourteen Points
9. The Washington Conference
10. The Japanese Exclusion Act
11. The Kellogg-Briand Pact
12. The Stimson Doctrine
13. Roosevelt's Quarantine Speech
14. The Atlantic Charter
15. The Connally Resolution
16. The Yalta Agreements
17. The Potsdam Agreement
18. United States Proposals for the International Control of Atomic Power
19. The Truman Doctrine
20. The Marshall Plan
21. The Point Four Program
22. The North Atlantic Treaty
23. American-Japanese Defense Pact
24. Atoms for Peace: Eisenhower's Proposal to the United Nations
25. The Formosa Resolution
26. The Eisenhower Doctrine
27. Alliance for Progress
28. Kennedy's Grand Design
29. Treaty on the Peaceful Uses of Outer Space

TEST DESCRIPTION REVIEW FORM

I. Objective

1. Do you know who the learner is and what curriculum area is covered?

2. From the behavior stated in the objective, imagine how it would be tested. Then, skip down to the sample item and see if it matches one of those expectations.

3. Can you imagine in a rough form the kind of content that might be appropriate for teaching to the objective (e.g. the objective will focus on concept learning or acquisition of historical facts, etc.)

4. Skip to the content limits and see if the content reflects your expectations. If the objective gives absolutely no clue as to what might be appropriate content, it probably should be written more specifically.

II. Sample item

1. Do you know what you're supposed to do with the item? Are directions clear or is the format so familar that you know how to record your answer?

2. Is the content presented to student what you'd expect based upon a reading of the objective and the content limits specifications?

3. Are the answer choices actually assessing/requiring the skills specified in the objective and response criteria (if a constructed response item)?

4. Do answer choices in multiple choice items conform to the item generation rules specified in the response criteria section?

III. Content Limits

1. Can you plan a lesson based on the content specifications? Would you know which materials to choose and which you should not choose?

2. Think of the most extreme example of content that would fit the objective and that some teacher might select. Ask yourself/or the specification's author if this would be appropriate. If not, the author/or you must write a rule saying so. If so, it's fine the way it is.

3. Are there any guidelines for selecting content so that it is appropriate in readability, difficulty or conceptual load for the target group? Are these needed for this specification given the entry skills of the target population and the intent of the objective?

4. Is this section organized in a way that you understand the content selection rules and how they relate to each other?

IV. Response limits

1. Multiple choice: are there rules for writing appropriate distractors? Do these distractor rules relate to how students learn and what are realistic mistakes for someone who has not mastered the concepts/skills tested?

IV. Response Limits

2. Constructed response:

- Are there rules for what content needs to be included in the answer (if appropriate)?
- Are there rules for the structure and organization of the written/oral response?
- Are there rules dealing with the kinds of mistakes students make and how they should be counted; eg extraneous information, well-written but off-topice, mechanical errors, omitted information?
- Are there rules for scoring the above response criteria that allow you to make a valid judgment about the student's mastery of the objective?

3. Observation schedules:

- Do the items in the schedule reflect the content limits and skill specified in the objective?
- Does the type of response the rater is coding reflect a valid theory of the relationship between the observed behavior and mastery of the objective?
- Are there counting/tallying/scoring rules that allow you to summarize the observations in such a way that you can decide if the objective has been mastered?

BEST COPY AVAILABLE

APPENDIX 2

Sample Constructed Response Scoring Guidelines

BEST COPY AVAILABLE

Grade 9

Rubric for Fall Semester Composition - Autiobiographical Incident

Task: Student writes a multi-paragraph autobiographical essay which discusses a personal change brought about by a specific incident. The change may involve altered behavior, attitudes, or perceptions. In the essay, the student must describe the change and show how it is attributable to the incident. While the incident should be clear, the focus of the essay should rest with the changes it brought about, rather than with the incident itself.

5: Applies to papers that do everything well. Includes all criteria for a 3 paper and, in addition,

a) is informative and entertaining
b) gives the reader insights about the author's past and present values, attitudes, beliefs, personality, points of view, growth and maturity, etc.
c) adds drama to commonplace experiences and elevates the everyday occurrence to a new level of force and meaning
d) uses vivid, concrete language/imagery which provides the paper with unique style or "voice"
e) has an engaging beginning and a body which fully explains how the experience made a difference in the author's viewpoints, behavior, etc.
f) exhibits excellent proofreading/editing skills.

4: Meets all the criteria of a 3 paper, and has one or two characteristics of a 5 paper.

3: Meets all of the following criteria:

a) Content:
- provides at least a minimal description of the incident and shows some evidence of its effect on the author's beliefs, attitudes, personality, etc.
- uses predominantly appropriate word choice

b) Organization:
- has a distinct beginning, middle, and end of the narrative
- has a title

c) Mechanics:
- uses substantially correct sentence structure
- has substantial freedom from serious grammatical/mechanical errors
- has neat appearance: white, lined paper, blue/black ink, legible cursive/manuscript writing using upper and lower case letters
- is at least 300 words
- has at least three paragraphs.

2: Meets some but not all of the criteria of a 3 paper

or

contains mechanical/grammatical errors which interfere with the reading of the paper.

1: Fails to address the task as directed

or

is illegible.

Revised 7/83

APPENDIX 3

Sample Test Item Formats

BEST COPY AVAILABLE

VOCABULARY ITEMS

IDENTIFYING SYNONYMS

Destroy a nest
1) **discover**
2) **hide in**
3) **tear up**
4) **describe**

This item is asking students to identify words virtually in isolation.

Automobiles cost a lot to operate. Bicycles, on the other hand, are <u>economical</u> to maintain.

The context provided for these synonyms is a sentence

F thrifty
G exciting
H difficult
J exhausting

ANTONYMS

The <u>opposite</u> of *to deny* is *to* _____

A. reply
B. affirm
C. destine
D. annoy

Words are identified in isolation.

DEFINITIONS

The word _____ means *knives*.

A. scurry
B. knead
C. cutlery
D. institute

With a one-word definition, item is asking for a synonym.

IDENTIFYING TWO MEANINGS FOR ONE WORD

30 <u>a high card</u> and <u>a military pilot</u>

A ace
B king
C aviator
D navigator

BEST COPY AVAILABLE

RECOGNIZING CORRECT/INCORRECT CAPITALIZATION

1) "Who has seen the wind?"
2) is the first line of the poem
3) our class learned today.
4) (No mistakes)

Student identifies line on which error occurs. Correct answer does not indicate whether student can supply correct capitals . Incorrect answer does not show what kind of error student makes.

1) My friend had to pay a dime
2) for the Library Book he
3) kept at home too long.
4) (No mistakes)

The enthusiastic band (a) played "Auld lang Syne" (b) at midnight (c) on New Year's Eve. (d) No error (e)

Student identifies specific word needing capitalization. Can pinpoint types of capitalization errors.

Select the underlined word in the sentence below that should not be capitalized.

Our lockers are located on the north side of dolly madison hall.

a north
b dolly
c madison
d hall

As in item above, recognition is limited to specific words.

IDENTIFYING BEST OPTION FOR CAPITALIZATION

Read the passage and letter and look at the numbered, underlined parts. Choose the answer that shows the best capitalization and punctuation for each part.

Vickie was winning a game of cat's cradle for the first time, she (25) was so excited that she dropped a loop, and the string (26) fell to the floor. She said "let's (27) play another game!"

25 **F** time. she
G time. She
H time? She
J Time, She
K Best as it is

27 **F** said, "Let's
G said "Let's
H said, Let's
J said. Let's
K Best as it is

...ization/punctuation in context

SELECTING WORD THAT COMPLETES THE MEANING OF A SENTENCE

Because of their weak supports, many buildings fell during the earthquake. Our school remained standing because of its strong _____.

A balcony
B ceiling
C pavement
D foundation

Tests both word meaning and understanding of certain concepts connected with architecture.

SELECTING WORDS TO COMPLETE THE MEANING OF A PARGRAPH

The ship's sides creaked and groaned as it battled its way through the _____ (42) storm. I was _____ (43) that the crashing waves would burst through my cabin walls and wash me out to sea. The wind was _____ (44), snapping at the sails as though it were a huge sea monster attacking our vessel. I knew I had to escape from that _____ (45) ship before it broke in half.

A brisk
B noisy
C raging
D distant

F careful
G fearful
H impressed
J undisturbed

44 **A** silent
B patient
C vicious
D poisonous

45 **F** sunken
G doomed
H forgotten
J torpedoed

This item is essentially a Cloze test. The focus is not so much on word meaning as in comprehension of text, as measured by selection of appropriate words.

PUNCTUATION ITEMS

RECOGNITION THAT PUNCTUATION IS NEEDED

1) The place I liked best on
2) our trip last August
3) was Denver Colorado.
4) (No mistakes)

Student indicates line on which punctuation is needed or is done incorrectly. Does not ask student to identify which type is needed.

SELECTION OF CORRECTLY PUNCTUATED SENTENCE

Which one of the following sentences is punctuated correctly?

a I wonder how hard my math class will be?
b I wonder how hard my math class will be.
c I wonder how hard my math class will be,
d I wonder how hard my math class will be:

Existence of alternatives makes item somewhat "easier" to sight readers.

INSERTION OF PUNCTUATION INTO CORRECT PLACE

Please pass the chili sauce Charles wants to put it on his roast beef.
A B C D

INDICATION WHAT PUNCTUATION NEEDED

"My paper airplane," Bob said, won't fly."

A . B ? C ! D " E None

This is similar to the first item on this page, except that students must indicate what kind of punctuation is needed.

Select the punctuation to end the following sentence:

Please let me know when I will receive the tickets

a ?
b .
c !
d :

Similar to above with the exception that student is told where punctuation should be inserted.

USAGE/GRAMMAR ITEMS

IDENTIFYING USAGE ERRORS

1) The bus has went without you.
2) We played all day at the beach.
3) Our baby can sit up by himself.
4) (No mistakes)

Doesn't ask student to supply correct answer
May test several rules in one item.

IDENTIFYING CORRECT USAGE

Select the sentence in which the underlined word is used correctly.

a The locker assigned to us are hard to reach.
b Each of us want the other's locker.
c We know others who also needs to change lockers.
d The locker problem is not too difficult to solve.

Difficult to come up with good distractors.

In which sentence below is the underlined word used incorrectly?

a Maria helps me with my homework.
b Tom and Joe need lots of help with his homework.
c Susie always does her homework at school.
d All of us take our homework seriously.

Similar to first item except only one usage point is tested.

RECOGNIZING CORRECT APPLICATION OF GRAMMAR RULES

Read the four sentences. Find the one sentence that has the correct verb to go with its subject.

61 F Someone has taken all the books in my locker.
G Janine and Carmine has lost all their books, too.
H Here is the books that were taken.
J One of the books are still missing.

Student can't do if she/he doesn't know terminology of "subject" and "verb."

SELECTION OF CORRECT GRAMMATICAL FORM

Teresa was _____ to do our experiment than I.

Tests one point in a sentence. Distractors focus most often on frequently confused usages.

A anxious
B anxiously
C most anxious
D more anxious

Pepe and Marty fastened the logs together _____ (14) formed a raft. _____ (15) they pitched a tent on it. _____ (16) shaping a rudder from a long piece of lumber, they used poles to push the raft from the shore. _____ (17) they didn't hurry, they soon reached the mouth of the river.

14
F or
G but
H nor
J and

15
A When
B Then
C Where
D Therefore

16
F Until
G After
H Because
J Moreover

17
A Soon
B Whenever
C Although
D In addition

Resembles a Cloze reading test. Student must comprehend passage and relationships between concepts in order to select correct transitions. Focus here is on meaning rather than selection of "standard" or "correct" forms. Thus, the "correctness" of a word is not is usage or some rule-based application but its ability to complete the meaning of the paragraph.

SPELLING ITEMS

IDENTIFYING WORDS SPELLED CORRECTLY

The lost dog was _____.

Student must recognize correct variation of one word.

A identifyd

B identifyed

C identiffed

D identified

Choose the sentence with no misspelled words.

Student must pick out word from the sentence. Most resembles proof-reading.

a I am very active in school organizations.

b Thier are not many people whom I dislike.

c I am interested in studing business education.

d Even through my parents are strict with me, I feel secure.

IDENTIFYING WORDS SPELLED INCORRECTLY

1) **pilow**
2) **tells**
3) **walking**
4) **sets**
5) (No mistakes)

This type of item allows you to test more spelling words because students must know how to spell all words in the item instead of recognizing correct variatio of one word.

Which one of the following sentences contains a misspelled word?

Item doesn't tell you if student knows which word is misspelled.

a I might be comeing to visit you next summer.

b Mom, naturally, is a little worried about my going alone.

c Mom hasn't decided yet if I may go.

d I suppose I'll know for sure in a few weeks.

Select the underlined word in the sentence below that is spelled incorrectly.

We are writting this letter to suggest a more satisfactory way to issue lockers.

Focus is on individual word. Like first item in this section but has words "in context." Also resembles proof-reading.

a writting

b letter

c suggest

d satisfactory

TOPIC SENTENCE

_____. They were painted and carved columns of wood. The carvings traced important events of the tribe. The totem poles often stood as memorials to the dead.

A Carvings of animals and people can be found on totem poles.

B Totem poles were an important form of Indian art and history.

C The Indians often believed that one animal was sacred to a tribe.

D Black, red, white, and blue-green colors were used to paint totem poles.

Student must read entire paragraph to identify best sentence.

RELEVANT SUPPORT

Select the sentence that best supports the main idea stated below.

One mistake by a teammate caused the team to lose.

a Bob wore blue and white striped shorts.

b The other team lived in Lost Valley County.

c Bob dropped the baton and slowed our team's time.

d The finish line on the track was clearly painted.

DEVELOPMENT

Our last basketball game of the season was a real heartbreaker.

a Hendrix was spectacular on both defense and offense. He stole the ball several times and scored twenty-three points, making him the top scorer in the game.

b As a good will gesture, our coach started over to congratulate the other coach after the game. But the other coach was picked up by his jubilant team and carried away.

c The fourth quarter was a seesaw battle that ended in a tie, so the game went into overtime. Then, two of our best players fouled out, giving the other team enough shots from the foul line to win.

d No one expected our team to make it to the regional play-off anyway, so losing that last game to the Buffaloes was not surprising. However, we did make a showing.

Asks student to select best development in terms of given topic sentence. Students must understand the concept of "heartbreaker" and select sentences that relate to this idea.

PARAGRAPH COMPOSITION ITEMS

SENTENCE ORDER

[1]The patterns used in the cloth identify each clan.
[2]Clan members have a strong family spirit.
[3]As part of this spirit, clans have developed their own special clothing.
[4]In Scotland, family groups are called clans.

Student must "hold" sentences in head as he/she tests possible correct sequences.

A 1 — 4 — 3 — 2
B 2 — 3 — 1 — 4
C 3 — 2 — 4 — 1
D 4 — 2 — 3 — 1

UNITY/COHERENCE

Which sentence does not belong in a paragraph that begins with the following sentence?

I have many interests outside of school.

a I like singing in my church choir.
b I am quite tall for my age.
c Last summer I entered a swimming meet.
d I get satisfaction from doing volunteer work at the hospital.

Read the paragraph. Choose the sentence that does *not* belong in the paragraph.

1. Over 1500 miles long, the Great Wall of China was built entirely by hand. 2. Designed to keep out invaders, it was begun during the Ch'in period. 3. During the Han period, the Chinese invented paper. 4. It took more than 100 years to build the wall.

F Sentence 1
G Sentence 2
H Sentence 3
J Sentence 4

DICTION/WORD CHOICE

Choose the most specific word or words to complete the sentence.

Wendy's new sweater has a _____ collar.

A. turtleneck
B. large
C. very unusual
D. nice

May be testing student's ability to supply concrete description. Student must understand concept of "specific" to supply correct answer.

SENTENCE CORRECTNESS (FRAGMENTS/RUN-ONS)

Which one of the following should be separated into two sentences?

a I'm large but not so fat as my older sister.
b Dieting is out of the question eating is one of the things I do best.
c Sometimes I wear make-up, although my mother doesn't approve.
d I'm quite intelligent, but I don't work very hard.

Doesn't require that students know the vocabulary of sentences; e.g. fragment and run-on

SENTENCE REVISION

Watching the horror movie, the castle door slowly creaked open.

F Slowly creaking open, I watched the castle door in the horror movie.
G I watched the horror movie slowly creaking as the castle door opened.
H Watching the horror movie, I saw the castle door slowly creak open.
J Best as it is

Tests ability to identify improvements in sentences based on clarification of a meaning.

SENTENCE COMBINING

My friend won the election.
It was for class secretary.
He is a new student.

Tests ability to recognize the sentence that produces an accurate relationship between ideas in a set of simple sentences.

A Since he is a new student, my friend won the election for class secretary.
B He is a new student, my friend who won the election for class secretary.
C My friend, who is a new student, won the election for class secretary.
D My friend is the new class secretary after winning the election as a new student.

BEST COPY AVAILABLE

Examples of items in horizontal and vertical formats

$36 - 5 =$

21

29

30

31

None of these

$$\begin{array}{r} 14 \\ -\ 12 \\ \hline \end{array}$$

1

2

6

20

None of these

$3 + 414 + 6 =$

423

433

504

1314

None of these

$$\begin{array}{r} 376 \\ 502 \\ +\ 503 \\ \hline \end{array}$$

1301

1370

1371

1381

None of these

$$\begin{array}{r} 8 \\ \times 3 \\ \hline \end{array}$$

a 12

b 18

c 21

d 27

BEST COPY AVAILABLE

$9\overline{)72}$

Ⓐ 6
Ⓑ 7
Ⓒ 8
Ⓓ 12
Ⓔ NH

$16 \div 2 = \square$

6 (a) 7 (b) 8 (c) 9 (d) NH (e)

61. 296 ÷ 13

A. 22 R11
B. 22 R9
C. 23
D. Not given

Although most published tests express remainders as "R", many math teachers are uncomfortable with this and would rather that remainders be expressed as fractions or decimals. Choice of how remainder will be express should depend upon how students are <u>taught</u> to express remainders.

$\square \times 9 = 27$

7 (f) 5 (g) 4 (h) 3 (i) NH (j)

The "unknown" in computation problems can be located in any part of the problem; Students may also need to know how to supply missing information other than the "answer." However, in problems like these, if they test more than multiplication/addition/subtraction/division facts, they may be testing the operation inverse to the one specified by the item.

41

$5 \times \square = 35$

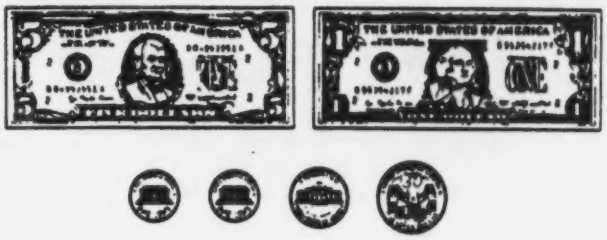
6 (a) 7 (b) 8 (c) 9 (d) NH (e)

42

$2 \times 84 = \square$

168 176 88 86 NH

33 **What is the value of these bills and coins?**



$5.82 $6.32 $6.37 $6.57 NH

Ⓐ Ⓑ Ⓒ Ⓓ Ⓔ

Which amount of money shown above equals $9.58?

A. 4
B. 2
C. 1
D. None of these

Money problems may use pictures of money rather than words.

Which container can hold the most water?

a cup

a pint

a quart

a gallon

You may want to test measurement concepts by asking students to understand the relationship between different units of measure. Anther item type similar to this is one that asks the student, "which unit of measure would you use to measure a pencil?"
(choices: centimeter, meter, kilometer)
or: centimeter, centigrade, gram)

There are 1000 meters in 1 kilometer. How many kilometers are there in 5000 meters?

5

15

50

55

There are many kinds of translations that can be made with measurement problems: within the English system
within the metric system
English to metric and visa-versa
The difficulty of these translations is increased if the "formula" is not provided.

44. **A Grant City map looks like graph paper. Avenues run north and south; numbered streets run east and west. Use the formula below to find the closest street to 229 Euclid Avenue.**

1. Drop the last figure of the address number.
2. Divide the remaining number by 2.
3. Add this number to the correct key number below. The new number is the nearest cross street.

Avenues	Key Numbers
Plymouth	Add 2
Euclid	Add 3
Midland	Add 5
Washington	Add 4

229 Euclid Avenue is located near—

A. 14th St.
B. 16th St.
C. 22nd St.
D. 6th St.

This problem asks students to follow directions. If they can read, they can apply the formula. No translation from symbols to words is required.

$$V = \frac{1}{3}\pi r^2 h$$

31. **If h = 6 cm and r = 3 cm, what is the volume of the figure above?**

A. 6π cu cm
B. 36π cu cm
C. 18π cu cm
D. None of these

This is an example of a symbolic formula.

15. **Given the percent formula, p = rb, where p = percentage, r = rate, and b = base, what is the percentage if the rate is 7% and the base is $2100?**

A. $130
B. $147
C. $140
D. None of these

Embedding this formula in a paragraph makes this problem somewhat difficult. Problem is one is substitution.

WORD PROBLEMS

John had 5 toy cars.
He bought 4 more cars.
He gave away 2 cars.
How many cars does he have now?

Straightforward word problem. Some word problems are longer and have "extraneous" information so that student is required to separate relevant from irrelevant material.

$30 off chest
FREEZER
249.96 regular
$279.96
9.0 cu. ft. chest freezer has a removable or lift-out basket. Walnut color trim. White.

Word problems may have picture stimuli. Often "real life" problems use this format.

6. If you buy the freezer shown in the ad above and pay for it in 12 equal monthly installments without interest charges, how much will each payment be?

A. $23.33
B. $22.83
C. $18.33
D. None of these

1 Jack had 15 apples. Then he gave away 7. Which number sentence can you use to find how many apples he has left?

Ⓐ 15+7=□ Ⓒ 15−7=□
Ⓑ 7+15=□ Ⓓ □−7=15
Ⓔ □−15=7

The problem could require that students understand how to "set up" or symbolize the problem rather than find a solution.

This is an attempt to test the student's problem solving ability.

2 Sue's family has 2 cats and 3 dogs as pets.

BEST COPY AVAILABLE

BEST COPY AVAILABLE

16 Louise and Tom have been playing. It is now 4 P.M. What else do you need to know to find how long they have been playing?

- ⊙ What time they have to be home
- ① How many children were playing
- ⊙ What game they were playing
- ○ What time they started
- ⊙ What time school ended

Asks for missing information. Attempts to test reasoning.

18 Alex wants to buy a book. Which statement is NOT needed to find how much more money he needs to buy the book?

- ① He had already saved 50¢.
- ⊙ The book costs $1.50.
- ⊙ He has been saving for 3 weeks.
- ① His mother gave him 10¢ more.
- ⊙ He earned 65¢ more.

Asks students to separate relevant from irrelevant information. Attempts to get at the problem solving process.

Susis wants to buy a bicycle costing $250. If she earns $3.75 in her after-school job and works 20 hours a week, how long will it take her to buy the bicycle.? In order to answer this question you must:

a. Divide and multiply
b. add and divide
c. multiply and subtract
d. multiply and divide

This type of problem can test problem solving strategies and order of operations concepts

BEST COPY AVAILABLE

FIGURE IDENTIFICATION ITEMS

Which of these is most like a cone?

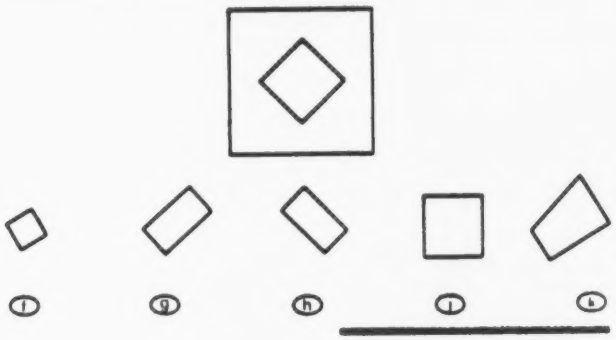
This item tests students to identify geometrical shaps in "everyday" form.

19 Which figure is NOT a triangle?

Requires students to identify positive and negative instances of the concept.

20 Which figure has the same size and shape as the figure in the box?

Figure is rotated and embedded



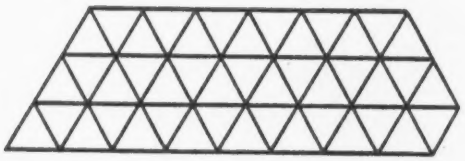


Figure is embedded.

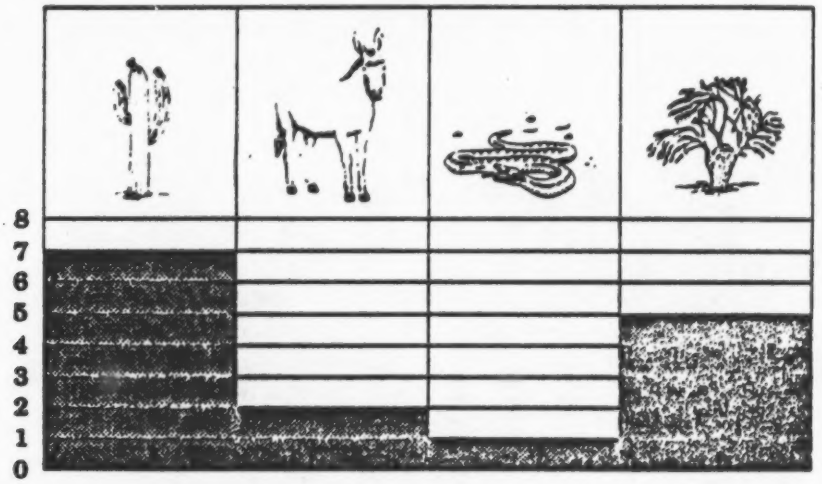
What shapes are repeated in the figure above?

A. Squares
B. Rectangles
C. Triangles
D. None of these

BEST COPY AVAILABLE

BEST COPY AVAILABLE

The graph shows pictures of four things that live near Gabriela's home. The numerals show how many of each kind of thing Gabriela saw when she took a nature walk.



Which of these did Gabriela see most often?

○

○

○

○

Item directs student how to read the graph.

BEAVERS' TREES

Pokey	
Brownie	
Silky	
Kitty	

Each stands for 1 tree.

Picture graph. Student should read "key" below graph, although not necessary for solving this problem.

4 How many trees did Pokey cut down?

BEST COPY AVAILABLE

AGES OF PUPILS IN IN OUR CLASS

Age (Years)	Number of Pupils
$7\frac{1}{2}$	/
8	///
$8\frac{1}{2}$	///// /////
9	///// ///// /
$9\frac{1}{2}$	///// /

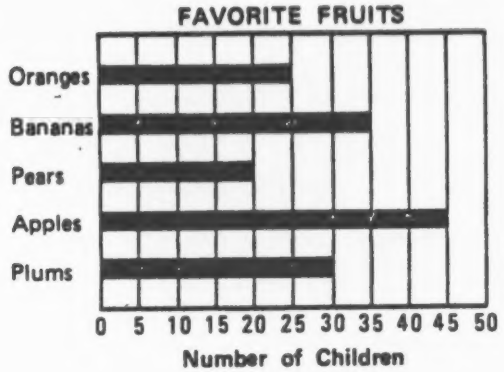
Each / means one pupil.

In which age group are there just 3 pupils?

$7\frac{1}{2}$ (f) 8 (g) $8\frac{1}{2}$ (h) 9 (i) $9\frac{1}{2}$ (k)

Tally graph--student needs to understand tallies.

The graph shows how many children chose different kinds of fruits as their favorites. Use this graph to answer questions 27, 28, and 29.



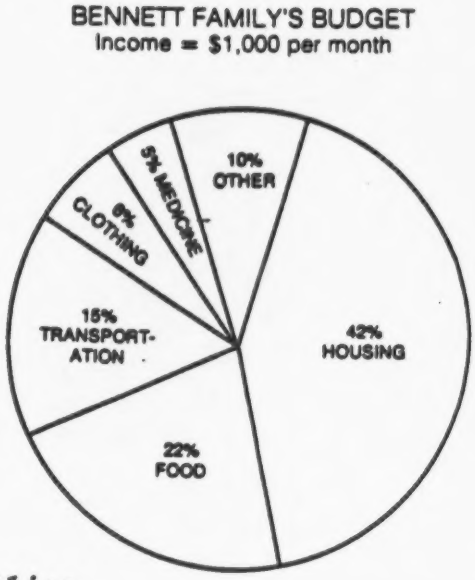
27 Which fruit was chosen by the fewest children?

ⓐ Oranges ⓒ Pears

ⓑ Bananas ⓓ Apples

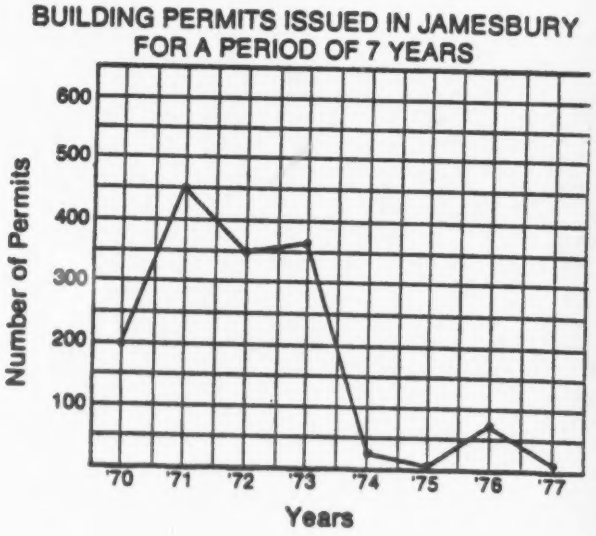
ⓔ Plums

Bar graph



. How much per month did the Bennett's spend on housing?

A. $42.00
B. $220.00
C. $420.00
D. None of these



89. In what year were the most building permit issued?

A. 1973
B. 1971
C. 1977
D. None of these

Line graph

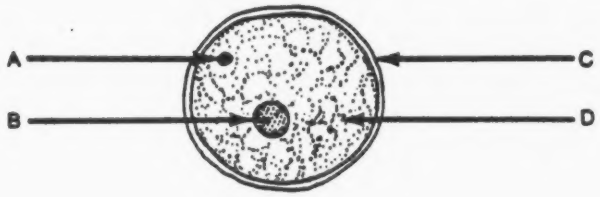
BEST COPY AVAILABLE

How do most bears spend the cold winter months?

Tests a fact.

- Eating fish
- Hunting for food
- Storing food
- Sleeping in caves

Concepts
Tests a definition.



Which is the nucleus of the cell?

a A
b B
c C
d D

A ball rolling down a hill has what kind of energy?

Concepts--example

f Kinetic
g Chemical
h Nuclear
j Radiant

What is true of many animals as they go into hibernation?

Generalizations/principles

a Their activity increases.
b Their rate of breathing increases.
c Their layers of fat become thicker.
d Their hearts beat faster.

Two nails were placed in water at the same time, and one nail rusted. The other nail did NOT rust because it probably –

- ⊙ had a different shape
- ⊙ was a bit larger
- ⊙ was a darker color
- ⊙ was made of a different material

Requires knowledge of principles of metal.

Sheep need vitamin B_{12} to be healthy. Some bacteria that live inside sheep make vitamin B_{12} for the sheep.

Which of these is true about bacteria?

F **They live only in sheep.**

G **They always cause disease.**

H **They need vitamin B_{12} to live.**

J **They can help as well as harm.**

Formal reasoning problem.

3 In an experiment, four groups of dogs are fed the diets shown below for 2 months.

Group	Diet
I	Wheat
II	Corn
III	Corn with three vitamins added
IV	Mixed diet of vegetables, milk, and meat

At the end of the experiment, the dogs in group IV are the healthiest. Which of the following is the broadest conclusion supported by the experiment?

- A Vitamins are not needed by dogs who eat corn.
- B Wheat is not good for dogs.
- C Corn is not good for dogs.
- D A varied diet is better than a diet of only one type of food.

Student needs to understand "broadest conclusion". Requires students to choose best answer according to criterion of generalizability.

Mary wants to learn more about what kind of rock she found. Which of these should she do?

Problem solving item set in context of using the "scientific method" or empirical model.

smell it

shake it

measure its size

test its hardness

In setting up an experiment to find out which evaporates faster, alcohol or water, you would need to control for all of the following except the—

Student must understand concept of "control".

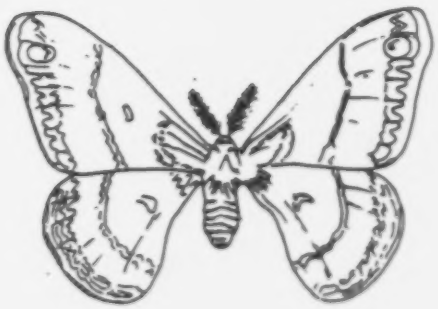
a freezing point of each liquid
b temperature of each liquid
c amount of each liquid
d exposed surface area of each liquid

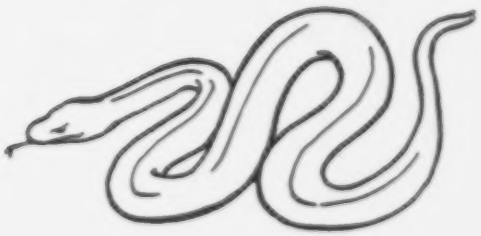
When the animal in the picture above becomes an adult, it will look like which of the following?

Unlike the majority of test items, asks for visual relationship rather than verbal.

A

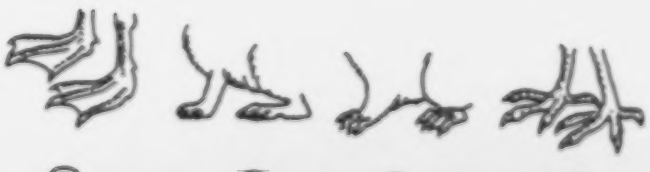
B

C 

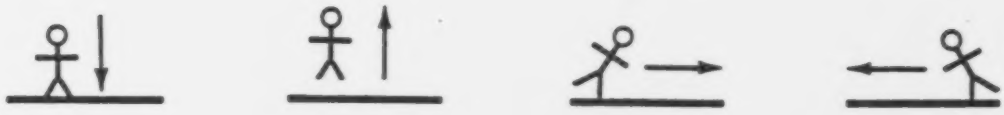
D 

1 Which picture shows the feet of an animal that spends much of its time in the water?

Answer choices can be visual.



Which picture shows the direction that gravity pulls everything on Earth?



A **B** **C** **D**

Requires student to symbolize verbal principle.

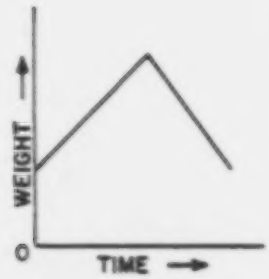
A certain baby loses weight because he cannot digest his mother's milk. The doctor finds something that the baby can digest and he gains weight. Which of the following graphs shows how his weight changes?

Asks for visualization/symbolization of verbal material. Translation task.

A

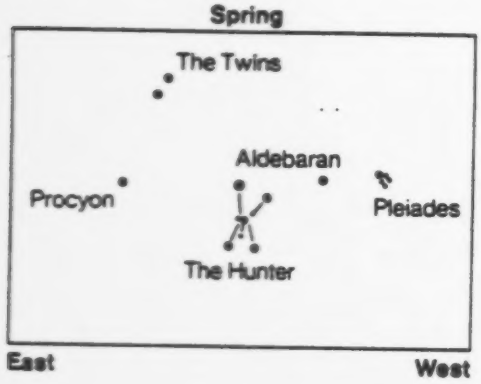
B

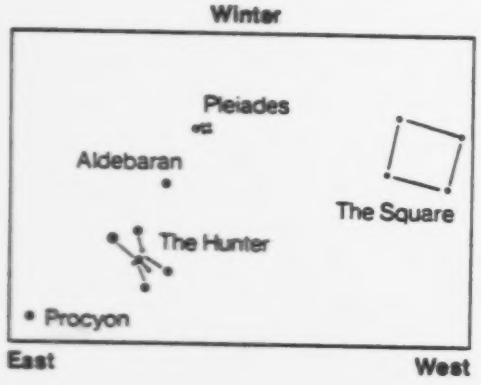
C



D

Many science items are testing students' ability to read the special material in science books. Newer items include maps of space and flow charts. You might even wish to test reading a programming manual.





Which star or group of stars can be seen only in winter?

- Procyon
- The Square
- Aldebaran
- The Hunter

How does The Hunter appear in the winter as compared to the spring?

- Farther to the east
- More spread out
- Higher in the sky
- Upside down

PLANT PRODUCTS

Plant Part	Products
fruits	jams, food products
saps	syrups, glues, paints
bark	dyes, drugs, flavors
logs	sawdust, lumber products

13 Fruit is used to make

- jelly
- paste
- aspirin
- maple syrup

PLANTS AND PLANT PARTS

PLANT	PLANT PART: Spores	Seeds	Flowers	Green Leaves
Fern	X			X
Daisy		X	X	X
Mushroom	X			
Pine Tree		X		X

27 Which plant has both spores and green leaves?

- fern
- daisy
- mushroom
- pine tree

Shows the variety of visual materials that students are required to use and understand in science. These are essentiall "reading" items.

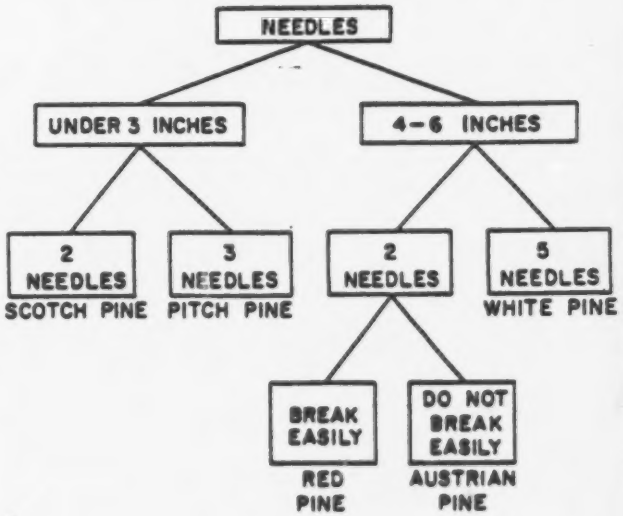
A science class kept a record of the time the sun came up and set each day. The times are shown on the table below.

DAY	DATE	SUNRISE	SUNSET
Monday	March 20	6:40 a.m.	5:48 p.m.
Tuesday	March 21	6:39 a.m.	5:49 p.m.
Wednesday	March 22	6:38 a.m.	5:50 p.m.
Thursday	March 23	6:37 a.m.	5:51 p.m.
Friday	March 24	6:36 a.m.	

29 What time did the sun probably set on Friday, March 24 ?

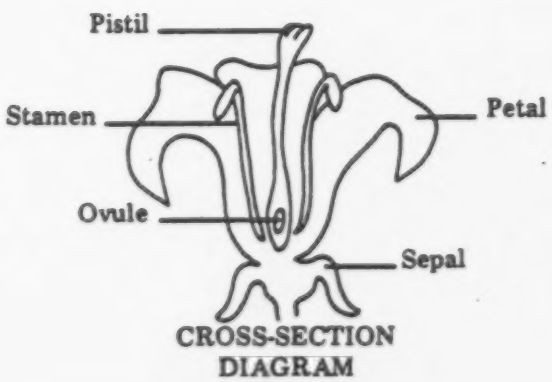
- ○ 5:52 p.m.
- ○ 5:53 p.m.
- ○ 5:54 p.m.
- ○ 5:55 p.m.

Questions 11-12 are based on the following diagram that describes needles on trees.

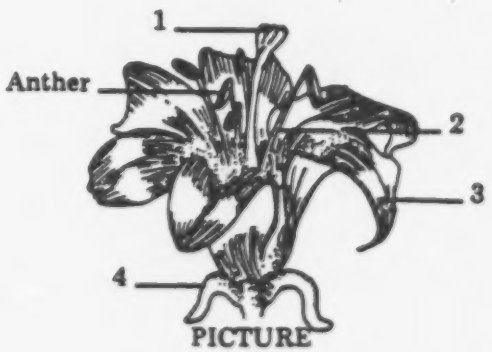


11 According to the diagram, if a tree has broad leaves, which of the following is a correct statement?

- A It is a white pine.
- B It is a Scotch pine.
- C It is a red pine.
- D It is none of the trees listed in the diagram.



CROSS-SECTION DIAGRAM



PICTURE

Compare the diagram and the picture. Which number in the picture identifies the sepal?

F 1

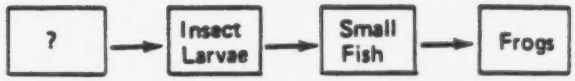
G 2

14 From the diagram and the picture, you can tell that an anther is part of the

A petal

B ovule

C stamen

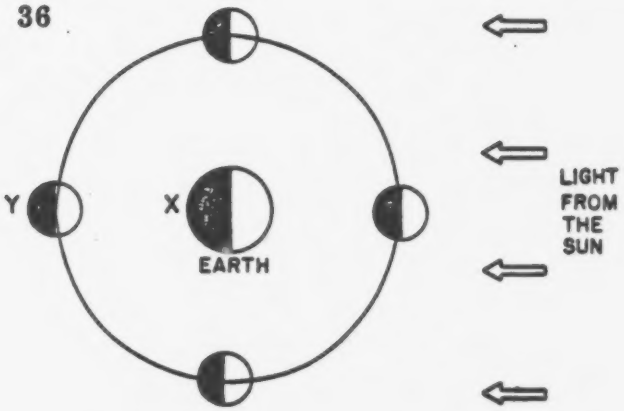


Analogy/reasoning problem with visual stimulus.

Which of the following could be the first link in this food chain?

f Fungi
g Algae
h Shrimp
j Fish eggs

36



Requires student to translate "flat" map into 3-D form.

The diagram above shows the Moon in four positions in its orbit around the Earth. The light parts of the Moon and the Earth show where the Sun is shining. If you are standing at point X on the Earth and the Moon is in position Y, how does the Moon look?

A

B

C

D

Scoring criteria would have to include definition of "effective." for last item.

Give two ways that you could legally prove that you took the leather coat to the store to be cleaned. Write your answers in the answer booklet.

You detect a strong odor of gas in the house when you come home late one night. The pilot light is off on the gas stove. The rest of the family is asleep. Which one of the following is the best action to take FIRST?

(1) Wake up your family.

(2) Call the gas company.

(3) Go to bed.

(4) Open the windows.

(5) Light the pilot light.

Gloria saw a want ad for a job as a florist's delivery driver. She needs the job. She has commercial driving experience and an excellent driving record. She applies for the job, but another person gets it. This person doesn't have any commercial driving experience. What is the most effective action for Gloria to take?

(1) Threaten to sue the manager for fraud.

(2) Inform the appropriate authorities that her rights have been violated.

(3) Threaten to tell the story to a friend who is a newspaper reporter.

(4) Ignore the whole thing. The other person probably should get the job anyway.

This NAEP exercise is an attempt to measure the Social Interaction domain of the NCSS Curriculum Guidelines. Again, criteria for each of the categories "helpful" to "harmful" would have to be established.

On this and the next two pages is part of a conversation that took place in a high school student committee meeting. Georgia is leading the committee and starts the conversation. How helpful or harmful are the responses of the other committee members in reaching the goal of the committee? Is each response "Very Helpful," "Helpful," "Harmful," or "Very Harmful?"

Georgia: "I guess you all know I was asked to be the head of this committee to plan a dance for our class. I hope we can make this the best dance we've ever had."

	Very Helpful	Helpful	Harmful	Very Harmful
A. Mary Alice: "Well, we never have any good dances. The chaperones always watch us like hawks."	(1)	(2)	(3)	(4)

BEST COPY AVAILABLE

SOCIAL STUDIES: READING SKILLS-

Tests student's ability to infer from social studies type text author viewpoint etc. In addition, student must understand such facts, concepts, information as the personalities and political viewpoints of the Founding Fathers, the New Jersey Plan and the Constitution.

"Manufacturing has positive benefits to offer society and therefore should be encouraged by the government."

Of the following, who was the strongest advocate of the viewpoint expressed in the quotation?

A Thomas Jefferson
B Benjamin Franklin
C Alexander Hamilton
D James Madison

Speaker III: You both miss the point in arguing about the distribution of powers between states and the central government. What our present system of government lacks is a chief executive and an effective system of national courts. There is danger in a concentration of power in one branch of government.

Speaker IV: I see nothing wrong with our present government. Congress has power to do the really necessary things. It has solved one of our major problems. Give it time and it will solve the others also. The only change which may be needed is a change in the unanimity rule.

26 Which speaker is closest to the position taken by the New Jersey Plan of Union?
A I B II C III D IV 26

27 The major problem to which Speaker IV refers is probably that of
A organizing the western territories.
B controlling Indian raids.
C establishing a common currency.
D making Britain live up to the Treaty of Paris. 2

28 The danger Speaker III mentions was a major reason for incorporating which of the following principles in the Constitution?
A Checks and balances
B Representative government
C Freedom of expression
D Due process of law

BEST COPY AVAILABLE

SOCIAL STUDIES: MULTIPLE CHOICE -TRUE/FALSE ITEM

An item format for true-false items.

Gail wants to buy her mother an electric appliance that will be safe for her mother to use. Does each of the following guarantee that the appliance meets minimum electrical safety standards?

	Does guarantee safety	Does not guarantee safety	I don't know.
A. A seal of approval from a well-known magazine	(1)	(2)	(3)
B. The salesperson's statement that the appliance is safe	(1)	(2)	(3)
C. An Underwriters Laboratories (UL) stamp or tag on the appliance	(1)	(2)	(3)
D. A 12 month warranty on the appliance	(1)	(2)	(3)

One of the main jobs of the wagon train leader years ago was to see that the people –

- ○ cared for the children
- ○ fed the horses
- ○ cooked the food
- ○ obeyed the rules

Tests fact.

41 A community is any group of people who –

- ○ look alike
- ○ shop in the same stores
- ○ have the same background
- ○ have common interests

Concept--definition

Houses, clothes, and languages are part of people's —

- f culture
- g art
- h religion
- j habits

Concept--examples

27 Which material s different from the other three?

- ○ Copper
- ○ Iron
- ○ Tin
- ○ Wood

Concept--negative and positive examples; classification

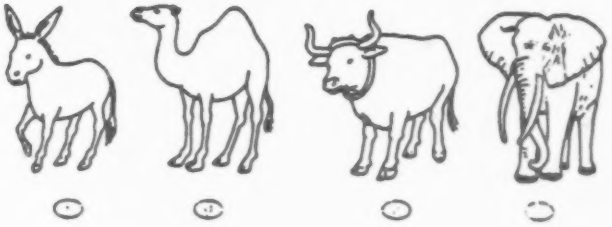
12 When it omes to food, all groups of people are like because all people –

- ○ mus eat to live
- ○ like he same kind of food
- ○ eat vith forks and knives

Generalization

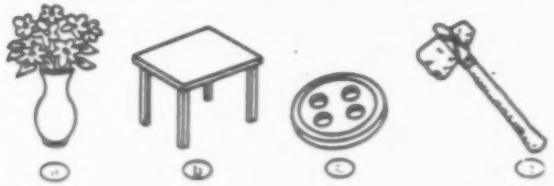
Facts, concepts, and generalizations may have visual as well as verbal representation.

32 **Which one of these animals is used for transportation by people who live high in the mountains?**



Fact

Which object is the best example of a tool?

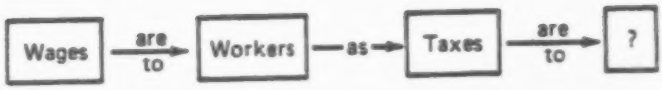


Concept--example

42 **Which form of transportation probably needs the FEWEST repairs?**

Generalization--number of moving parts=complex

23



① Clubs
② Sports
③ Government
④ Education

Measures analogy with visual prompt

Timeline

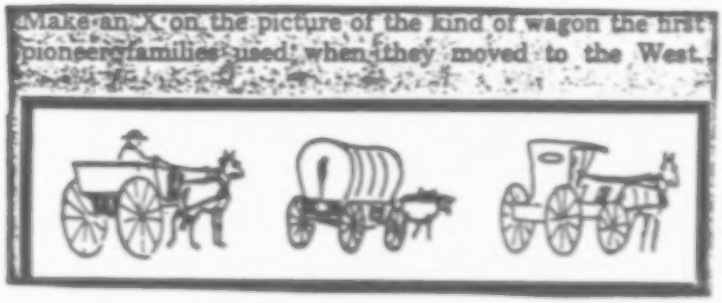
During which period did each of the following events take place?

17 Citizen Genêt threatened American neutrality.

18 Washington delivered his Farewell Address.

19 An undeclared naval war was carried on with France.

Make an X on the picture of the kind of wagon the first pioneer families used when they moved to the West.

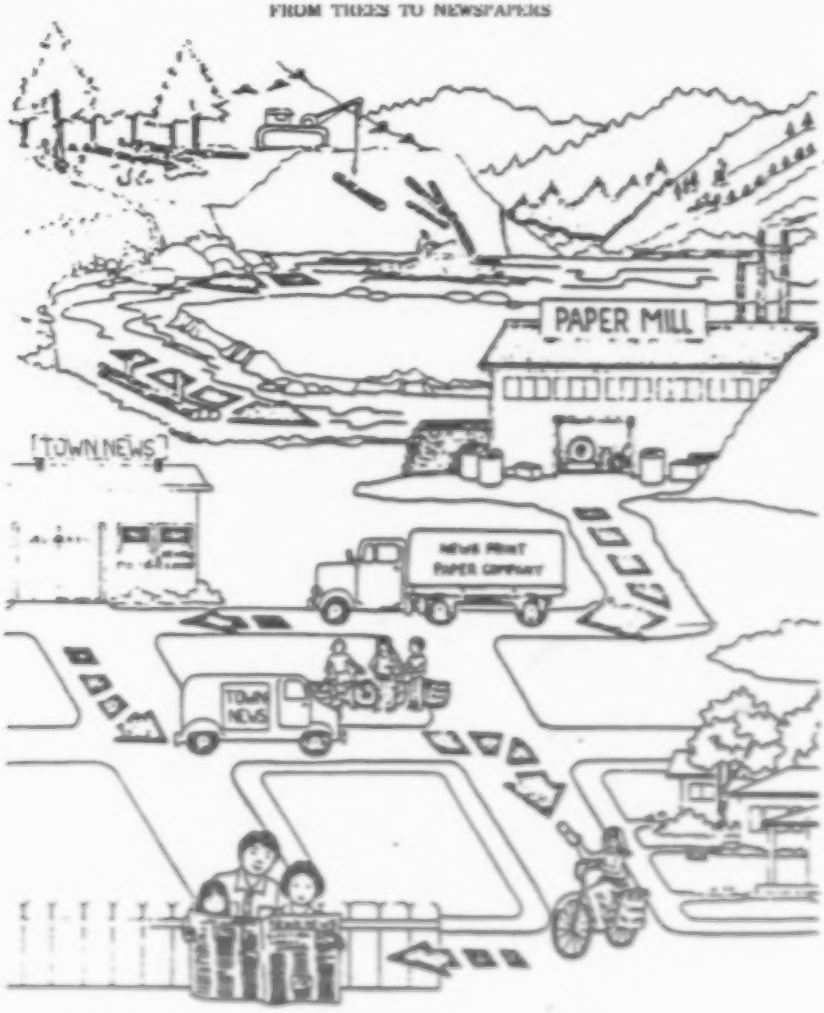


These objects were found buried in the ground. What do they tell about the people who used them?

① How they cared for their children
② What language they spoke
③ How they got their food
④ What kind of clothes they wore

Look at the picture. Then do Numbers 24 through 27.

FROM TREES TO NEWSPAPERS



Tests student's ability to read a diagram.

24 In the picture, what natural resource is used to move the logs to the mill?

- oil
- coal
- wind
- water

At the mill, the logs are

- made into paper
- burned for power
- used for building
- made into furniture

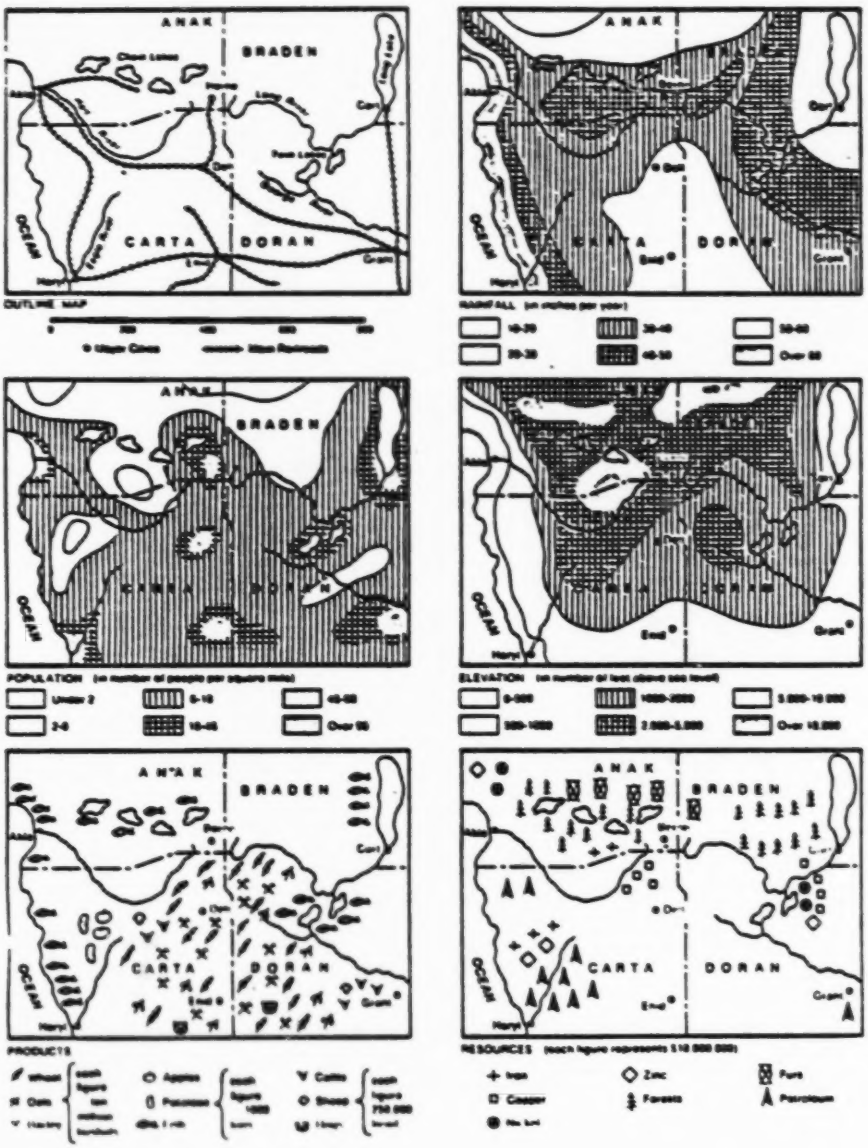
25 To deliver a newspaper, you need to know

- where your customers live
- how to cut down pine trees
- how to print the newspaper
- what is printed in the newspaper

27 Newspapers help people on election day by

- writing about sports
- reminding them to vote
- printing the weather report
- telling them where to [illegible]

BEST COPY AVAILABLE



The questions accompanying these maps often ask students to combine information from one or more maps in order to answer the question.

About how many head of cattle are sold each year in Carta?

1) 250,000
2) 500,000
3) 750,000
4) 1,000,000

Which city would be most likely to have an oil refinery?

1) Akla 2) Dell 3) Grant 4) Haryl

What might account for the sparsely populated area on and near the border between Anak and Carta?

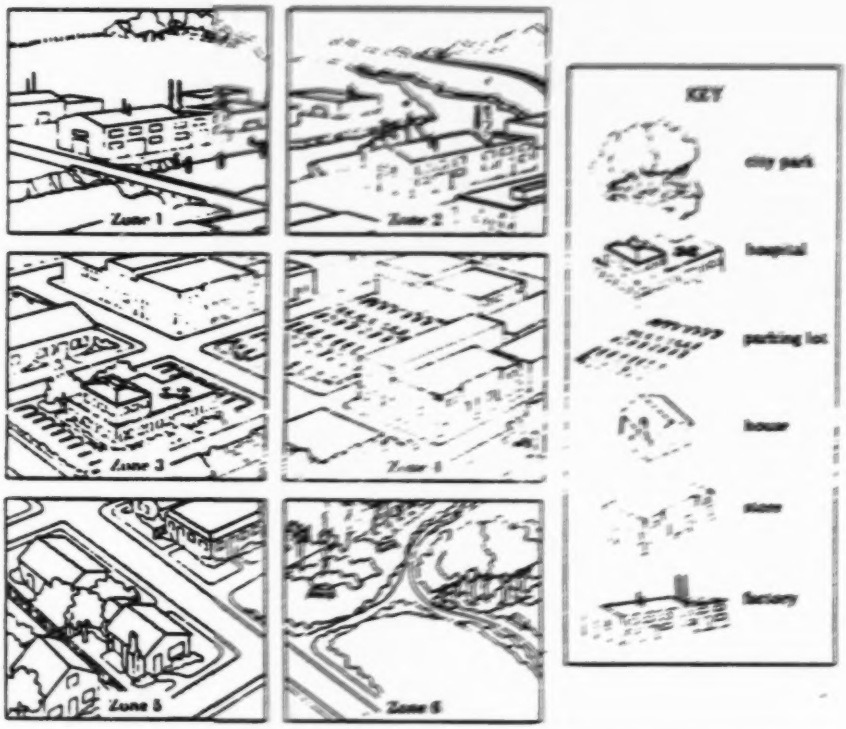
1) It is located in high mountains.
2) It has very little rainfall.
3) It is heavily forested.
4) There are no railroads serving the area.

Which of these might be the average rainfall per year in the major grain-producing areas?

BEST COPY AVAILABLE

A nearly three-dimensional map; useful for younger students.

Look at the picture below. It shows part of a city. Use the picture and the key to answer Numbers 13 through 15.



13 In what zone is the hospital?

3 55

4 66

14 In what zone would you probably find most people's homes?

2 55

3 66

15 In what zone would it be best to have a school picnic?

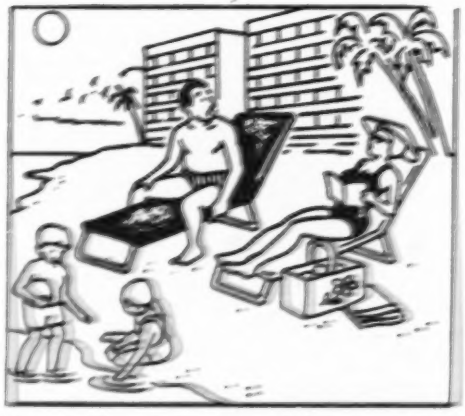
11 55

44 66

BEST COPY AVAILABLE

Relationships between place and event (integrative concepts).

8 The Whisky Rebellion took place in area
A 4. B 5. C 6. D 7.



"Mimi, I just figured out that every time a wave breaks it costs us 166 cents."

Cartoons

The cartoon above shows that—

A. Mimi is worried about the children

B. the beach is disappointing

C. the man did not leave his worries at home

D. the children don't like the water

BEST COPY AVAILABLE